BAYESIAN SEQUENTIAL METHODS FOR BINOMIAL AND MULTINOMIAL SELECTION PROBLEMS

by

SAAD ABED MADHI

A thesis submitted for the degree of Doctor of Philosophy at the University of Keele

Department of Mathematics,
University of Keele,
March, 1986.

TABLE OF CONTENTS

page

ACKNOWLEDGEMENTS

I am deeply indebted to my supervisor, Dr. P. W. Jones who has been a constant source of help, guidance and inspiration during the course of this work.

I would like to thank the Ministry of Higher Education in Iraq for the scholarship I received which helped make this study possible.

Special thanks are due to members of the Computer Centre of Keele University, particularly Dr. P. G. Collis and Dr. J. Galletly who have been consistently cooperative and helpful. Grateful acknowledgement is given to Dr. A. Mahendrasingam, Physics Department, for useful comments to solve computational problems on Computer CYBER 205.

I am very grateful to Mrs. O. Brindley for her patience and masterly typing of the thesis.

Finally, I wish to express my sincere appreciation to my family for their remarkable encouragement and support during my studies and my wife for her patience and moral support.

ABSTRACT

The thesis deals with binomial and multinomial sequential selection problems. Optimal sequential sampling schemes are derived by using dynamic programming in conjunction with loss functions and sampling costs and to minimize expected sample sizes. Comparisons are carried out with sampling schemes where observations are taken in blocks or groups and with fixed sample size procedures.

Several suboptimal designs are suggested and numerical comparisons are made under several performance characteristics which are obtained exactly and by using Monte Carlo simulation. The performance of the procedures is studied when the parameters are fixed and where they are generated from particular prior distributions.

CHAPTER 1

GENERAL INTRODUCTION

In many real-life situations, one is often faced with the problem of selecting the best (the term best is assumed to be well-defined before experiment) among several alternatives or ranking them according to their performance. For example, we may be interested in choosing the best of several drugs (treatments) or choosing best candidate from several alternatives. The statistical techniques by which these problems can be solved are known as ranking and selection procedures.

In this thesis we treat two selection problems concerning Binomial amd Multinomial distributions. Most of the previous work on these two problems seems to be limited to the classical or indifference zone approach (Bechhofer (1954)) which can be described as follows. A probability level P^* and critical difference Δ^* (called P^*, Δ^*-condition) where, $0 \leq P^*, \Delta^* \leq 1$, are chosen such that the probability of selecting the best alternative is at least as large as the preassigned number P^* whenever Δ, the difference between the value of the largest parameter and the value of the parameter next to the largest, is equal to or greater than the preassigned number Δ^*.

However, it appears that the restriction to the classical approach is unreal because of the lack of connection between the value of Δ^* and the true parameters of interest. Since Δ^* is to be specified by the experimenter and the probability of correct selection is guaranteed to be at least P^* only when $\Delta \geqslant \Delta^*$, in fact the true distance is unknown and there is no knowledge concerning the true probability of correct selection. Having regard to this difficulty, alternative approaches to the solution of these problems have to be considered. Bechhofer and Kulkarni (1981) proposed some alternative procedures which do not use the (P^*, Δ^*-condition).

In this thesis we attempt to apply Bayesian statistical decision theory which leads to a quite different approach to the selection problem as the concepts of loss of taking a certain decision when particular values of the parameters of interest are true, the cost of sampling and some prior information about the parameters of the underlying distributions are involved. Some mathematical results are given and those obtained by large scale Monte Carlo simulation.

Throughout this thesis we shall generally assume the following conditions:

1. There is prior knowledge regarding the parameters of interest.
2. The procedures are truncated (closed) where there exists an upper bound of number of observations that can be carried out until a decision is taken. Evidently, the procedure will terminate with probability one. The fact that these procedures are closed increases their potential for use in real-life applications.
3. The outcomes of the observations are independent and the probabilities for these outcomes remain constant from

observation to observation (stationary).

Furthermore, since reaching a decision as quickly as possible is desirable, it seems sensible to employ sequential techniques to achieve the aim. Different sequential sampling rules have been adopted. The observations are either taken sequentially one at a time or a group at a time and then sampling is stopped and a decision is taken, or sampling continues. The main properties of a sequential procedure are that the sample size required to terminate the procedure is a random variable since it depends on the results of the observations and they are economical in that a decision may be reached earlier by sequential procedure than by that using a fixed sample size.

The following rules have to be specified for any sequential procedure:

(a) Sampling rule.

The sampling rule prescribes which observations are taken from which populations. We consider different sampling rules such as one at a time or a group at a time.

(b) Stopping rule.

The stopping rule prescribes when sampling should be terminated. At this stage one population is chosen as the better.

(c) Termination rule.

A decision is made at the time when sampling is terminated.

As we mentioned earlier, this thesis consists of two selection problems, Binomial and Multinomial. The first problem represents the largest portion of the thesis and some optimal and suboptimal Bayesian sequential (including group sequential and fixed sample size) designs are presented for choosing the better of two Binomial populations. In these cases rules (a), (b) and (c) must be specified and hence they are problems in sequential design. The second problem deals with Bayesian sequential schemes for choosing the largest multinomial cell and briefly for choosing the ordering of the cells.

The plan of the thesis can be summarized as follows. Chapter 2 to Chapter 8 are devoted to the problem of selecting the better of two Binomial populations.

Chapter 2 is an introduction to the Binomial selection problem where the problem is stated and review of previous work on it as well as relevant definitions and notation are presented.

Chapter 3 discusses the Bayesian decision theoretic formulation of the problem. Using this formulation along with the dynamic programming technique introduced by Bellman (1957), a fully optimal sequential scheme has been constructed. The performance of this procedure has been investigated in terms of Bayes risk. The effect of changing the loss constants, sampling costs and prior information on the Bayes risk are also discussed. A good deal of attention is devoted to the problem of optimal sample size, including optimal sample size for the sequential scheme, using risks.

In Chapter 4, two optimal schemes, namely group sequential and fixed sample size schemes are proposed and investigated. The determination of the sample size and the effect of group size are also considered. Risk efficiencies for both schemes with respect to fully sequential scheme are calculated.

Some suboptimal schemes and their risks have been developed in Chapter 5. The determination of the optimal sample size is considered and the relative efficiencies of these schemes are also calculated.

In Chapter 6 Monte Carlo simulation studies have been carried out to investigate the performance of fully sequential, group sequential and fixed sample size schemes in terms of other performance measures such as the probability of correct selection and expected sample sizes.

An extensive simulation study which was performed to investigate the performance of the suboptimal schemes is reported in Chapter 7. Different sampling rules used in conjunction with Bechhofer and Kulkarni (1981) stopping rule with some modifications have been investigated. Numerical comparisons between the proposed stopping rule and the Bechhofer and Kultarni stopping rule are given.

Chapter 8 contains optimal and some suboptimal sequential designs for reducing the number of observations on the poorer population which is very important in the design of clinical trials.

In Chapter 9 we have developed Bayesian suboptimal schemes (single sequential, group sequential and fixed sample size) for the Multinomial selection problem. Here the goal is to select the largest cell probability in Multinomial distribution. Some relevant definitions and notation have been introduced. The methods of generating variates from the

Dirichlet distribution are developed. Some numerical comparisons with Kulkarni (1981) procedure are presented.

Some supplementary investigations and suggestions for further work are given in Chapter 10.

The appendices contain listings of the computer programs which have been used to produce the numerical part of this study. They may also be useful as source listings for users wanting programs for studying other rules.

CHAPTER 2

INTRODUCTION TO THE BINOMIAL SELECTION PROBLEM

2.0 Summary

In section 2.1 the problem of selecting the better of two Binomial populations is stated. Examples of applications of the problem are given.

Section 2.2 contains an historical review.

Section 2.3 gives some relevant concepts and definitions, on which the construction of the selection procedures based, are given.

2.1 The statement of the problem

Suppose that π_i $(i = 1, 2)$ are two Binomial populations. The quality of the i^{th} population is characterized by a real-valued parameter p_i, the unknown probability of a success in a single trial from population i, where $0 \leqslant p_i \leqslant 1$ $(i = 1, 2)$. The problem is to select the better of these Binomial populations on the basis of a sequence of observations. The better population is defined to be the one with the higher probability of success. The ranked success probabilities are denoted by $p_{[1]} \leqslant p_{[2]}$, where $p_{[1]} = \min(p_1, p_2)$ and $p_{[2]} = \max(p_1, p_2)$. The values of $p_{[j]}$ $(j = 1, 2)$ are assumed to be unknown to us. Moreover, we do not know which population is associated with $p_{[2]}$. Observations may be obtained sequentially, singly or in groups of constant size. Our goal is to design selection procedures that enable us to select the population associated with $p_{[2]}$; thus we have a two-decision problem.

The statistical formulation as stated above is typical of many well-known practical problems encountered in many situations in real-life. Two examples of fields of applications are presented below.

(a) Medical applications

Suppose that a stream of patients are to be allocated to one of two medical treatments and suppose that the result of each trial is known before the next patient is allocated the treatment, if it is assumed that the response is dichtomous with probability p_i that the treatment i results in a success or cure then the schemes proposed will suggest an allocation based on the single observation sequential sample. In practice group sequential designs (Pocock (1977)) would perhaps be the more useful. Furthermore, the experimenter may have to take account of ethical considerations in the allocation of the treatments, it is possible to argue that the objective in this case is to maximise the number of cures leading to the two armed bandit problem (Berry (1972)) rather than the two decision problem given above.

(b) Industrial applications

Consider a firm producing a particular product. Let M_1 and M_2 are two industrial production methods, making the same product. An observation from M_i means producing an item using the production method M_i ($i = 1, 2$). All items are being categorized simply as effective or defective. The effectiveness of a production method is evaluated in terms of the proportion of effective units produced by that production method. Here p_i is the unknown probability of an item being effective when it is produced using the production method M_i $(i = 1, 2)$. The aim is to select the production method that has higher probability of producing an effective item. Therefore, the problem of comparing two production methods is eventually a problem of comparing two Binomial populations in terms of their single-trial success probabilities.

Further examples on the situations where the Binomial model applies and the problem of practical interest is to select the better of two Binomial populations may be found in Gibbons, Olkin and Sobel (1977).

The following experimental conditions should be met:

1. The observations (trials) produced by each population are independent of each other.
2. p_1 and p_2, the probabilities of success are constant during the experiment.

A discussion of these conditions have been given in Büringer et al. (1980).

2.2 Review of literature

In the last two decades there has been an increasing interest in the development of selection procedures to solve the problem of selecting the better of two Binomial populations. This interest has stemmed from the potential applicability of these procedures in medical trials and related fields of applications.

Sobel and Huyett (1957) is considered as a fundamental paper in Binomial selection studies. In this paper they proposed a single sampling procedure in which an equal number of observations n^* are taken from each population and the population having the most successes is selected as the better population with ties broken by randomization. They employed the idea of indifference zone approach which was developed by Bechhofer (1954) to solve the problem of selection in Normal populations. This classical approach requires that the probability of making a correct selection is greater than or equal to some preassigned value, P^*, where in Binomial selection problem the true difference between the largest and the second largest in p-values is larger than or equal to another preassigned number, Δ^*. Formally

$$P(CS) \geq P^*, \tag{2.2.1}$$

whenever

$$p_{[2]} - p_{[1]} \geq \Delta^*, \qquad 0 < \Delta^* < 1, \ \tfrac{1}{2} < P^* < 1, \tag{2.2.2}$$

where 'CS' (for correct selection) denotes the final selection of a population with probability of success $p_{[2]}$.

With the condition above, called P^*, Δ^*-condition, we will be at least $100P^*$ percent sure of selecting the better parameter whenever the better parameter $p_{[2]}$ is at least Δ^* better than the second $p_{[1]}$. The P(CS) is minimized when

$$p_{[2]} - p_{[1]} = \Delta^*,$$

this is called the least favourable configuration (LFC) of the population parameters p_1, p_2. The value of n^* is then chosen to guarantee (2.2.1) when the parameter values in the least favourable configuration.

The problem of allocating (assigning) observations among patients in clinical trials has been investigated using other approaches by many authors. Armitage (1975) developed closed sequential procedures. Anscombe (1963), Colton (1963) and Canner (1970) used loss functions to produce a decision-theoretic approach using a fully sequential procedure. More recently Pocock (1977) has developed a group sequential design for clinical trials in which the data are analysed at less frequent intervals and which may lead to an early decision, or stopping of a clinical trial, if large treatment differences are observed.

Other workers using the indifference zone approach are Taylor and David (1962) who discussed a multistage procedure for this problem. Paulson (1967) who proposed an open sequential procedure which permitted the elimination of 'non-contending' population and Bechhofer, Kiefer and Sobel (1968) who proposed an open sequential procedure employing a vector at a time sampling rule (VT). The application of the play the winner sampling rule (PWR) to the problem of allocating observations among treatments appeared first in Zelen (1969).

Later a great deal of attention has been paid to the sequential procedures for this problem using different sampling rules such as PWR and VT sampling rules. Kiefer and Weiss (1971), Hoel (1972), Nebenzahl and Sobel (1972), Berry and Sobel (1973), Fushimi (1973), Kiefer and Weiss (1974), Simon et al. (1975), Sobel and Weiss (1970) and Tamhane (1985) proposed and studied closed sequential procedures for selecting the better of two Binomial populations. Procedures for selecting the best population of $k \geq 2$ Binomial populations were considered by Sobel and Weiss (1972), Hoel and Sobel (1972) and Hoel, Sobel and Weiss (1975a).

At the same time, another approach to the Binomial selection problem has been suggested in the classical framework; it is known as the subset selection approach. Here the goal is to select a subset containing the best population with a preassigned probability P^*. This approach is useful for the situation when we have very large number of populations and the procedures require more observations than are available. Therefore it is desirable to select a subset consisting of the best for further extensive investigation. Gupta, Huyett and Sobel (1957), Gupta and Huang (1976) are among those who studied this problem using this approach. Goel and Rubin (1977) gave a general Bayesian decision theoretic approach for selecting a subset containing the best population.

The main difference between the subset selection approach and the indifference zone approach is that in the subset selection we have no indifference zone and the least-favourable configuration is simply the worst configuration with all parameters are equal (for details see Gibbons, Olkin and Sobel (1977, 1979).

As has been pointed out in Chapter 1, the disadvantages of the classical approach is that it is difficult to decide on the values of P^* and Δ^* in advance. The approach also ignores the existence of any prior information about the parameter of interest.

Although the literature on Binomial selection problems is large, the literature using Bayesian approach to solve the problems is rather scarce. Important contributions were made by Bland and Bratcher (1968), Bratcher and Bland (1975), who developed Bayesian fixed sample size procedures to solve the problem of ranking Binomial probabilities where more than two populations are compared.

Recently, Bechhofer and Kultarni (1981) proposed a very interesting closed sequential procedure avoiding the (P^*, Δ^*-condition) of the indifference zone approach. In a subsequent paper, Bechhofer and Kulkarni (1982) gave exact numerical results for the performance characteristics of the procedures given in Bechhofer and Kulkarni (1981). Bechhofer and Frisardi (1982) investigated these procedures employing Monte Carlo simulation. They have also been discussed in Jennison (1983, 1984) and Kulkarni and Jennison (1984), Kulkarni and Kulkarni (1986) have studied them using Bayesian approach.

A two-decision problem for a single Binomial population using optimal sequential sampling is given in Lindley and Barnett (1965).

It seems worthwhile at this stage to mention a similar problem namely the 'two Armed-Bandit problem'. Here the goal is to allocate N observations, one at a time, between two Bernoulli processes so as to maximize the expected number of successes. This problem has been investigated by many authors such as Berry (1972), Wanhrenberger et al. (1977), Jones (1974, 1975), Fabius and Van Zwet (1970), Poloniecki (1979) and Meeter (1975).

A good review of literature of many proposed procedures with particular reference to adaptive sampling for clinical trials, is contained in Hoel, Sobel and Weiss (1975b) and a complete overview is contained in Gupta and Panchapakesan (1979). Dudewicz and Koo (1982) contains a comprehensive bibliography on statistical selection and ranking procedures.

In this thesis a Bayesian approach is adopted to the Binomial selection problem and optimal and suboptimal procedures are studied. Their performance is evaluated by calculating their Bayes risk and other criteria in some cases using Monte Carlo simulation.

2.3 Some basic ideas and concepts

This section presents some concepts and ideas which are useful in constructing the proposed section procedures.

1. Bayesian approach

Prior to the experimentation we may have some information about p_1 and p_2, encapsulated in a form of $\pi(p_i)$, called the prior probability function of p_i. It reflects our personal beliefs about the unknown parameter p_i. From the experimentation we can gain some information by observing the random variable X with probability density function $f(x|p_i)$. Let $\pi(p_i|x)$ denote the posterior probability density function of p_i, then by Bayes theorem

$$\pi(p_i|x) \propto \pi(p_i)\ f(x|p_i) \qquad (i = 1,\ 2).$$

It should be pointed out that the prior probability density and posterior probability density are relative terms. If data arrives in many stages, singly or in groups, the posterior beliefs derived from the previous stage forms the prior beliefs for the next stage. In fact the actual choice of $\pi(p_i)$ depends upon the statistician and the information and experience available to him at the time of doing the experiment. A prior which contains no information about p_i is called noninformative prior or vague prior. Mathematical and computational difficulties may arise from using some prior distributions. A reasonable method of overcoming these difficulties is to use a particular class of prior distributions, discussed by Wetherill (1961) and said to be closed under a given sampling distribution yielding the observations. This means that the form of the posterior distribution and prior distribution are identical, for example the Beta distribution is closed under sampling with respect to binomially distributed observations. Alternatively, this class of priors has been termed as natural conjugate priors in Raiffa and Schlaifer (1968).

In Bayesian statistical decision theory, there are two important components. The first component is the prior information about the parameters of interest. The second component is the loss function which depends on the parameters of interest and the decisions taken by the statistician. The method of solution is to compute the posterior expected loss and then select the decision that minimizes it. These are called the Bayes risk and Bayes decision respectively, the latter is also called the optimal decision.

2. Dynamic programming

Backward induction has been used in the literature as a computational technique for finding the optimal sequential procedures for different statistical decision problems. Bellman (1957) introduced alternative term to backward induction, called it dynamic programming and showed how it could be used to solve multistage decision processes. The general ideas of dynamic programming and its applications can be found in Simpson (1961). The linkage between dynamic programming and decision theory was given in Lindley (1961). Wetherill (1961), Lindley and Barnett (1965), Freeman (1970, 1972), EL-Sayyad and Freeman (1973) and Jones (1974, 1975) among others employed dynamic programming allied with the concept of truncation, that is the maximum number of observations is fixed in advance, to construct some optimal sequential procedures.

The dynamic programming technique is particularly important in multistage processes where decisions are taken sequentially and where they are not independent, in that decisions taken earlier will affect decisions made later. Consequently in order to find the best decision at the present time, it is necessary to know the best decision in the future and the only way to know that is to work backwards from the optimal future decisions back to the origin. Our procedures are developed in Chapters 3 and 4, using this technique.

CHAPTER 3

THE FULLY SEQUENTIAL SCHEME FOR SELECTING THE BETTER OF TWO BINOMIAL POPULATIONS

3.0 Introduction and summary

In this chapter we deal with the Bayesian sequential decision approach for the construction of a procedure for selecting the better of two Binomial populations. With the specifications of prior distribution on the unknown parameters and the specification of loss function we can formally write down the dynamic programming equations for the Bayesian sequential decision procedures.

A computer program has been written, given in appendix (3.1) to calculate the exact values of risks for the above procedure. There are two main difficulties in implementing such a program. The first is the computer time and the second is the computer storage.

The structure of this chapter is as follows.

In section 3.1 the two-decision Binomial selection problem is formulated.

In section 3.2 the stopping risks of various loss functions are given.

In section 3.3 some monotonicity properties of the stopping risks are discussed.

In section 3.4 the selection procedure, OPT_1, is constructed using the dynamic programming technique in conjunction with decision formulation given in section 3.1 and the stopping risks given in section 3.2.

In section 3.5 some properties of OPT_1 are given.

The schemes with specific loss constants, sampling costs and prior information are investigated in section 3.6.

In section 3.7, we discuss the problem of determining the optimal maximum sample size using risks.

3.1 Bayesian decision-theoretic formulation

Consider the two Binomial populations π_1 and π_2 with p_1 and p_2 as their unknown success probabilities for a single trial respectively.

Now, consider the following two-decision problem with decisions

$$D_1 : p_1 \leqslant p_2,$$

and

$$D_2 : p_1 > p_2. \tag{3.1.1}$$

Suppose the losses in making decisions D_1 and D_2 are given as follows:

$$L_1(D_1; p_1, p_2) = \begin{cases} 0 & \text{if } p_1 \leqslant p_2 \\ K_1W(p_1, p_2) & \text{if } p_1 > p_2 \end{cases} \tag{3.1.2}$$

and

$$L_2(D_2; p_1, p_2) = \begin{cases} K_2W(p_1, p_2) & \text{if } p_1 \leqslant p_2 \\ 0 & \text{if } p_1 > p_2 \end{cases} \tag{3.1.3}$$

where K_1, K_2 are positive constants giving losses in terms of cost, $W(p_1, p_2)$ is a function of p_1 and p_2. These loss functions with their stopping risks will be discussed in detail in the next section.

The Bayesian approach requires that we specify a prior probability density function $\pi(p_i)$, expressing our beliefs about p_i before we obtain the data. As we mentioned in section 2.3 from a mathematical point of view, it would be clearly very convenient if p_i is assigned a prior distribution which is a member of a family of distributions closed under Binomial sampling or as a member of the conjugate family. The conjugate family in this case is the family of Beta distributions. Accordingly, let p_i is assigned Beta prior distribution with parameters a_i and b_i, $Be(a_i, b_i)$. The normalized density function (Raiffa and Schlaifer (1961)) is given by

$$\pi(p_i) = \frac{p_i^{a_i-1}(1-p_i)^{b_i-a_i-1}}{\beta(a_i, b_i-a_i)},$$

$$0 \leqslant p_i \leqslant 1,\ 1 \leqslant a_i < b_i,\ i = 1, 2, \tag{3.1.4}$$

where $\beta(a_i, b_i - a_i)$ is the complete beta function. It is also assumed that p_i are a priori independent. The parameters a_i and b_i need not be integer. However it is convenient if from this point, we assume that a_i and b_i are integers so that we can replace the gamma functions by the factoral terms in our formulation of the schemes.

If our prior beliefs cannot be presented in a form of (3.1.4), the analysis which follows will not be applicable and the form of the posterior probability density function will have to be calculated directly from Bayes theorem.

The assessment of prior distributions is not pursued here, we assume that the statistician has already chosen his prior distribution, details on this may be found in Winkler (1972), Raiffa and Schlaifer (1961).

In addition to the prior information, we obtain some sample information from the population π_i (i = 1, 2). In doing so, we assume that we observe the number of successes c_i, obtained in d_i trials giving probability function

$$f(c_i | p_i, d_i) = \binom{d_i}{c_i} p_i^{c_i} (1 - p_i)^{d_i - c_i},$$

$$c_i \in \{0, \ldots, d_i\}. \tag{3.1.5}$$

The posterior probability density function is derived from the prior probability function and the assumed sampling model by means of Bayes theorem mentioned earlier.

$$\pi(p_i | r_i, n_i) = \frac{f(c_i | p_i, n_i)\ \pi(p_i)}{\int_{p_i} f(c_i | p_i, n_i)\ \pi(p_i)\ dp_i} \tag{3.1.6}$$

where $r_i = a_i + c_i$, $n_i = b_i + d_i$, $i = 1, 2$.

If the sample size d_i taken from population π_i is large, then the actual choice of prior parameters (a_i, b_i) has little effect on the posterior density function which can be well approximated by a Beta probability density function with parameters c_i and d_i. In this case it is sufficient to take the uniform prior distribution $\pi(p_i) = 1$, to express our vague knowledge about the parameters of interest.

As the Beta family is conjugate with the Binomial sampling, it is unnecessary to revise a Beta prior distribution on the basis of a sample from a Bernoulli process using Bayes theorem. Given the prior distribution and the sample results, we need simply note that

$$r_i = a_i + c_i \qquad \text{and} \qquad n_i = b_i + d_i$$

are the parameters of the posterior Beta density function.

3.2 The stopping risks S_1, S_2 for various loss functions

In this section, we derive the stopping risks of making decisions D_1 and D_2 for various loss functions. The stopping risk (the posterior expected loss) of the terminal decision D_1 when the posterior distribution for p_i has parameter (r_i, n_i) or alternatively when the sample path has reached (r_1, n_1, r_2, n_2) from the origin (a_1, b_1, a_2, b_2), denoted by $S_1(r_1, n_1, r_2, n_2)$, can be found as follows.

$$S_1(r_1, n_1, r_2, n_2) = E[L_1(D_1; p_1, p_2)]_{\pi(p_1, p_2 | r_1, n_1, r_2, n_2)} \qquad (3.2.1)$$

(where the subscript, $\pi(p_1, p_2|r_1, n_1, r_2, n_2)$, on the expectation sign is the joint posterior of p_1 and p_2 with respect to which the expectation is being performed), with S_2 similarly defined. The forms of $S_1(r_1, n_1, r_2, n_2)$ and $S_2(r_1, n_1, r_2, n_2)$ will be derived for the following loss functions.

1. Linear loss function

Suppose the losses in making decision D_1 and D_2 are linear in $|p_1 - p_2|$ then $W(p_1, p_2)$ in (3.1.2) and (3.1.3) has the form

$$W(p_1, p_2) = |p_1 - p_2|. \qquad (3.2.2)$$

Now,

$$S_1(r_1, n_1, r_2, n_2) = \int_0^1 \int_0^{p_1} K_1(p_1 - p_2)$$

$$\pi(p_1, p_2 | r_1, n_1, r_2, n_2)\, dp_2\, dp_1 \qquad (3.2.3)$$

$$= \frac{K_1(n_1 - 1)!\,(n_2 - 1)!}{(r_1 - 1)!(r_2 - 1)!(n_1 - r_1 - 1)!(n_2 - r_2 - 1)!}$$

$$\left[\int_0^1 \int_0^{p_1} p_2^{\,r_2-1}(1 - p_2)^{n_2-r_2-1}\right.$$

$$p_1^{\,r_1}(1 - p_1)^{n_1-r_1-1}\, dp_2\, dp_1$$

$$- \int_0^1 \int_0^{p_1} p_2^{\,r_2}(1 - p_2)^{n_2-r_2-1}$$

$$\left. p_1^{\,r_1-1}(1 - p_1)^{n_1-r_1-1}\, dp_2\, dp_1\right]$$

(since p_1, p_2 are independent).

Using integration by parts, we get

$$\int_0^{p_1} p_2^{\,r_2-1}(1 - p_2)^{n_2 - r_2 - 1}\, dp_2 = \sum_{j=r_2}^{n_2-1} \frac{(r_2 - 1)!\,(n_2 - r_2 - 1)!}{j!(n_2 - 1 - j)!}$$

$$p_1^{\,j}(1 - p_1)^{n_2-1-j}$$

and

$$\int_0^{p_1} p_2^{\,r_2}(1 - p_2)^{n_2-r_2-1}\, dp_2 = \sum_{j=r_2}^{n_2-1} \frac{r_2!\ (n_2 - r_2 - 1)!}{(j + 1)!\ (n_2 - 1 - j)!}$$

$$p_1^{\,j+1}(1 - p_1)^{n_2-1-j}$$

provided $r_2 > 0,\ n_2 - r_2 > 0.$

Hence,

$$S_1(r_1,\ n_1,\ r_2,\ n_2) = \frac{K_1(n_1 - 1)!\ (n_2 - 1)!}{(r_1 - 1)!\ (n_1 - r_1 - 1)!}$$

$$\left[\sum_{j=r_2}^{n_2-1} \left\{\frac{1}{j!\ (n_2 - 1 - j)!}\right.\right.$$

$$\int_0^1 p_1^{\,r_1+j}(1 - p_1)^{n_1+n_2-r_1-j-2}\, dp_1$$

$$- \frac{r_2}{(j + 1)!\ (n_2 - 1 - j)!}$$

$$\left.\left.\int_0^1 p_1^{\,r_1+j}(1 - p_1)^{n_1+n_2-r_1-j-2}\, dp_1\right\}\right]$$

with

$$\int_0^1 p_1^{\,r_1+j}(1 - p_1)^{n_1+n_2-r_1-j-2}\, dp_1 = \beta(r_1 + j + 1,\ n_1 + n_2 - r_1 - j - 1).$$

Therefore,

$$S_1(r_1, n_1, r_2, n_2) = \frac{K_1(n_1 - 1)!\ (n_2 - 1)!}{(r_1 - 1)!\ (n_1 - r_1 - 1)!}$$

$$\left[\sum_{j=r_2}^{n_2-1} \frac{(j + 1 - r_2)\ \beta(r_1 + j + 1, n_1 + n_2 - j - r_1 - 1)}{(j + 1)!\ (n_2 - 1 - j)!}\right] \tag{3.2.4}$$

$$= K_1 \sum_{j=r_2}^{n_2-1} (j + 1 - r_2)\ G_1(j) \tag{3.2.5}$$

where,

$G_1(j) = \dfrac{\beta(r_1 + j + 1, n_1 + n_2 - j - r_1 - 1)}{\beta(r_1, n_1 - r_1)\ \beta(j, n_2 - j)\ j(j + 1)}$, and $\beta(.,\ .)$ is a Beta function.

From which it follows that the stopping risk of the terminal decision D_2, denoted by $S_2(r_1, n_1, r_2, n_2)$, is given by

$$S_2(r_1, n_1, r_2, n_2) = \frac{K_2(n_1 - 1)!\ (n_2 - 1)!}{(r_2 - 1)!\ (n_2 - r_2 - 1)!}$$

$$\left[\sum_{j=r_1}^{n_1-1} \frac{(j + 1 - r_1)\ \beta(r_2 + j + 1, n_1 + n_2 - j - r_2 - 1)}{(j + 1)!\ (n_1 - 1 - j)!}\right]$$

$$= K_2 \sum_{j=r_1}^{n_1-1} (j + 1 - r_1)\ G_2(j) \tag{3.2.6}$$

where,

$$G_2(j) = \frac{\beta(r_2 + j + 1,\ n_1 + n_2 - j - r_2 - 1)}{\beta(r_2,\ n_2 - r_2)\ \beta(j,\ n_1 - j)\ j(j + 1)}.$$

2. Quadratic loss function

In this type of loss functions, suppose $W(p_1, p_2)$ in (3.1.2) and (3.1.3) is quadratic in $|p_1 - p_2|$ and has the form

$$W(p_1, p_2) = |p_1 - p_2|^2. \qquad (3.2.7)$$

Hence,

$$S_1(r_1, n_1, r_2, n_2) = \int_0^1 \int_0^{p_1} K_1 (p_1 - p_2)^2$$

$$\pi(p_1, p_2 | r_1, n_1, r_2, n_2)\ dp_2\ dp_1 \qquad (3.2.8)$$

$$= \frac{K_1 (n_1 - 1)!\ (n_2 - 1)!}{(r_1 - 1)!\ (r_2 - 1)!\ (n_1 - r_1 - 1)!\ (n_2 - r_2 - 1)!}$$

$$\int_0^1 \int_0^{p_1} (p_1^2 - 2p_1p_2 + p_2^2)\ p_1^{r_1 - 1}$$

$$(1 - p_1)^{n_1 - r_1 - 1}\ p_2^{r_2 - 1}\ (1 - p_2)^{n_2 - r_2 - 1}\ dp_2\ dp_1$$

$$(3.2.9)$$

$$= \frac{K_1 (n_1 - 1)!\ (n_2 - 1)!}{(r_1 - 1)!\ (r_2 - 1)!\ (n_1 - r_1 - 1)!\ (n_2 - r_2 - 1)!}$$

$$[I_1 - 2I_2 + I_3] \qquad (3.2.10)$$

where,

$$I_1 = \int_0^1 \int_0^{p_1} p_1^2 p_1^{r_1-1} (1 - p_1)^{n_1-r_1-1} p_2^{r_2-1} (1 - p_2)^{n_2-r_2-1} dp_2 \, dp_1$$

$$I_2 = \int_0^1 \int_0^{p_1} p_1 p_2 p_1^{r_1-1} (1 - p_1)^{n_1-r_1-1} p_2^{r_2-1} (1 - p_2)^{n_2-r_2-1} dp_2 \, dp_1$$

$$I_3 = \int_0^1 \int_0^{p_1} p_2^2 p_1^{r_1-1} (1 - p_1)^{n_1-r_1-1} p_2^{r_2-1} (1 - p_2)^{n_2-r_2-1} dp_2 \, dp_1 .$$

Now, using integration by parts we get

$$I_1 = \int_0^1 \left[\int_0^{p_1} p_2^{r_2-1} (1 - p_2)^{n_2-r_2-1} dp_2 \right] p_1^{r_1+1} (1 - p_1)^{n_1-r_1-1} dp_1$$

$$= \int_0^1 \left[\sum_{j=r_2}^{n_2-1} \frac{(r_2 - 1)! \, (n_2 - r_2 - 1)!}{j! \, (n_2 - 1 - j)!} p_1^{j} (1 - p_1)^{n_2-j-1} \right]$$

$$p_1^{r_1+1} (1 - p_1)^{n_1-r_1-1} dp_1$$

$$= \sum_{j=r_2}^{n_2-1} \frac{(r_2 - 1)! \, (n_2 - r_2 - 1)!}{j! \, (n_2 - 1 - j)!} \int_0^1 p_1^{j+r_1+1}$$

$$(1 - p_1)^{n_1+n_2-r_1-j-2} dp_1$$

$$= \sum_{j=r_2}^{n_2-1} \frac{(r_2 - 1)! \, (n_2 - r_2 - 1)!}{j! \, (n_2 - 1 - j)!} \beta(r_1 + j + 2, n_1 + n_2 - r_1 - j - 1)$$

$$I_2 = \int_0^1 \left[\int_0^{p_1} p_2^{r_2} (1 - p_2)^{n_2-r_2-1} dp_2 \right] p_1^{r_1} (1 - p_1)^{n_1-r_1-1} dp_1$$

$$= \int_0^1 \left[\sum_{j=r_2+1}^{n_2} \frac{r_2(n_2 - r_2 - 1)!}{j!\,(n_2 - j)!} p_1^{\,j}(1 - p_1)^{n_2-j} \right] p_1^{\,r_1}$$

$$(1 - p_1)^{n_1-r_1-1}\, dp_1$$

$$= \int_0^1 \left[\sum_{j=r_2}^{n_2-1} \frac{r_2!\,(n_2 - r_2 - 1)!}{(j + 1)!\,(n_2 - j - 1)!} p_1^{\,j+1}(1 - p_1)^{n_2-j-1} \right]$$

$$p_1^{\,r_1}(1 - p_1)^{n_1-r_1-1}\, dp_1$$

$$= \sum_{j=r_2}^{n_2-1} \frac{r_2!\,(n_2 - r_2 - 1)!}{(j + 1)!\,(n_2 - j - 1)!} \beta(j + r_1 + 2,\ n_1 + n_2 - r_1 - j - 1)$$

$$I_3 = \int_0^1 \left[\int_0^{p_1} p_2^{\,r_2+1}(1 - p_2)^{n_2-r_2-1}\, dp_2 \right] p_1^{\,r_1-1}(1 - p_1)^{n_1-r_1-1}\, dp_1$$

$$= \int_0^1 \left[\sum_{j=r_2+2}^{n_2+1} \frac{(r_2 + 1)!\,(n_2 - r_2 - 1)!}{j!\,(n_2 - j + 1)!} p_1^{\,j}(1 - p_1)^{n_2-j+1} \right]$$

$$p_1^{\,r_1-1}(1 - p_1)^{n_1-r_1-1}\, dp_1$$

$$= \int_0^1 \left[\sum_{j=r_2}^{n_2-1} \frac{(r_2 + 1)!\,(n_2 - r_2 - 1)!}{(j + 2)!\,(n_2 - j - 1)!} p_1^{\,j+2}(1 - p_1)^{n_2-j-1} \right]$$

$$p_1^{\,r_1-1}(1 - p_1)^{n_1-r_1-1}\, dp_1$$

$$= \sum_{j=r_2}^{n_2-1} \frac{(r_2 + 1)!\ (n_2 - r_2 - 1)!}{(j + 2)!\ (n_2 - j - 1)!} \int_0^1 p_1^{\,j+r_1+1}$$

$$(1 - p_1)^{n_1+n_2-j-r_1-2}\ dp_1$$

$$= \sum_{j=r_2}^{n_2-1} \frac{(r_2 + 1)!\ (n_2 - r_2 - 1)!}{(j + 2)!\ (n_2 - j - 1)!}\ \beta(j + r_1 + 2,\ n_1 + n_2 - j - r_1 - 1)$$

Hence, equation (3.2.10) becomes

$$S_1(r_1,\ n_1,\ r_2,\ n_2) = \frac{K_1(n_1 - 1)!\ (n_2 - 1)!}{(r_1 - 1)!\ (n_1 - r_1 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{\beta(j + r_1 + 2,\ n_1 + n_2 - j - r_1 - 1)}{(j + 2)!\ (n_2 - j - 1)!}$$

$$\{(j + 2)(j + 1) - 2r_2(j + 2) + (r_2 + 1)r_2\}$$

$$= \frac{K_1(n_1 - 1)!\ (n_2 - 1)!}{(r_1 - 1)!\ (n_1 - r_1 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \left\{ \frac{(r_1 + j + 1)!\ (n_1 + n_2 - j - r_1 - 2)!}{(j + 2)!\ (n_2 - j - 1)!\ (n_1 + n_2)!} \right.$$

$$\left. \times\ (j^2 + 3j - 2jr_2 + r_2^{\,2} - 3r_2 + 2) \right\} \tag{3.2.11}$$

or

$$S_1(r_1, n_1, r_2, n_2) = K_1 \sum_{j=r_2}^{n_2-1} [G_1(j)(r_1 + j + 1)$$

$$(j^2 + 3j - 2r_2 j + r_2{}^2 - 3r_2 + 2)/(j + 2)(n_1 + n_2)]$$

(3.2.12)

where $G_1(j)$ as defined in (3.2.5).

From which it follows that

$$S_2(r_1, n_1, r_2, n_2) = \frac{K_2(n_1 - 1)!\ (n_2 - 1)!}{(r_2 - 1)!\ (n_2 - r_2 - 1)!}$$

$$\sum_{j=r_1}^{n_1-1} \left\{ \frac{(r_2 + j + 1)!\ (n_1 + n_2 - j - r_2 - 2)!}{(j + 2)!\ (n_1 - j - 1)!\ (n_1 + n_2)!} \right.$$

$$\left. \times\ (j^2 + 3j - 2jr_1 + r_1{}^2 - 3r_1 + 2) \right\}$$

(3.2.13)

or

$$S_2(r_1, n_1, r_2, n_2) = K_2 \sum_{j=r_1}^{n_1-1} [G_2(j)(r_2 + j + 1)$$

$$(j^2 + 3j - 2jr_1 + r_1{}^2 - 3r_1 + 2)/(j + 2)(n_1 + n_2)]$$

(3.2.14)

where $G_2(j)$ as defined in (3.2.6).

3. Constant loss function

If the losses are constant, then

$$W(p_1, p_2) = 1. \tag{3.2.15}$$

Therefore,

$$S_1(r_1, n_1, r_2, n_2) = \int_0^1 \int_0^{p_1} K_1 \, \pi(p_1, p_2 | r_1, n_1, r_2, n_2) \, dp_1 \, dp_2 \tag{3.2.16}$$

$$= \frac{K_1 (n_1 - 1)! \; (n_2 - 1)!}{(r_1 - 1)! \; (n_1 - r_1 - 1)! \; (r_2 - 1)! \; (n_2 - r_2 - 1)!}$$

$$\int_0^1 \left[\int_0^{p_1} p_2^{\,r_2 - 1} (1 - p_2)^{n_2 - r_2 - 1} \, dp_2 \right] p_1^{\,r_1 - 1}$$

$$(1 - p_1)^{n_1 - r_1 - 1} \, dp_1$$

provided $r_2 > 0$, $n_2 - r_2 > 0$

$$= \frac{K_1 (n_1 - 1)! \; (n_2 - 1)!}{(r_1 - 1)! \; (n_1 - r_1 - 1)!}$$

$$\int_0^1 \left[\sum_{j=r_2}^{n_2 - 1} \frac{1}{j! \; (n_2 - 1 - j)!} \, p_1^{\,r_1 - 1 + j} \right.$$

$$\left. (1 - p_1)^{n_1 + n_2 - r_1 - j - 2} \right] dp_1$$

$$= \frac{K_1 (n_1 - 1)! \; (n_2 - 1)!}{(r_1 - 1)! \; (n_1 - r_1 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{1}{j!\ (n_2 - 1 - j)!} \left[\int_0^1 p_1^{\ r_1-1+j} (1 - p_1)^{n_1+n_2-r_1-j-2}\ dp_1\right]$$

$$= \frac{K_1 (n_1 - 1)!\ (n_2 - 1)!}{(r_1 - 1)!\ (n_1 - r_1 - 1)!} \sum_{j=r_2}^{n_2-1} \frac{\beta(r_1 + j,\ n_1 + n_2 - r_1 - j - 1)}{j!\ (n_2 - 1 - j)!} \tag{3.2.17}$$

we can put (3.2.17) in the form

$$S_1(r_1, n_1, r_2, n_2) = K_1 \sum_{j=r_2}^{n_2-1} \frac{(j + 1)(n_1 + n_2 - 1) G_1(j)}{(r_1 + j)} \tag{3.2.18}$$

where $G_1(j)$ is defined as before.

From (3.2.16) and (3.2.17) and (3.2.18) it follows

$$S_2(r_1, n_1, r_2, n_2) = \frac{K_2 (n_1 - 1)!\ (n_2 - 1)!}{(r_2 - 1)!\ (n_2 - r_2 - 1)!} \sum_{j=r_1}^{n_1-1} \frac{\beta(r_2 + j,\ n_1 + n_2 - r_2 - j - 1)}{j!\ (n_1 - 1 - j)!} \tag{3.2.19}$$

or

$$S_2(r_1, n_1, r_2, n_2) = K_2 \sum_{j=r_1}^{n_1-1} \frac{(j + 1)(n_1 + n_2 - 1)G_2(j)}{(r_2 + j)} \tag{3.2.20}$$

where $G_2(j)$ has the form given in (3.2.6).

In the case of constant losses, note that

$$S_1(r_1, n_1, r_2, n_2) = K_1 \, P(p_1 > p_2 | r_1, n_1, r_2, n_2)$$

and

$$S_2(r_1, n_1, r_2, n_2) = K_2 \, P(p_1 \leq p_2 | r_1, n_1, r_2, n_2).$$

3.3 Monotonicity properties of the stopping risk

In this section we present some properties of the stopping risk with some numerical work. These monotonicity properties occur as a result of varying the values of r_1, n_1, r_2, n_2 and have been investigated under the linear loss function through the following cases.

(1) $S_1(r_1 + 1, n_1, r_2, n_2) - S_1(r_1, n_1, r_2, n_2)$

(2) $S_1(r_1, n_1 + 1, r_2, n_2) - S_1(r_1, n_1, r_2, n_2)$

(3) $S_1(r_1 + 1, n_1 + 1, r_2, n_2) - S_1(r_1, n_1, r_2, n_2)$

(4) $S_1(r_1, n_1, r_2 + 1, n_2) - S_1(r_1, n_1, r_2, n_2)$

(5) $S_1(r_1, n_1, r_2, n_2 + 1) - S_1(r_1, n_1, r_2, n_2)$

(6) $S_1(r_1, n_1, r_2 + 1, n_2 + 1) - S_1(r_1, n_1, r_2, n_2)$

No assumption is made about whether some points are reachable by any sample path.

At the point (r_1, n_1, r_2, n_2),

$$S_1(r_1, n_1, r_2, n_2) = \frac{K_1(n_1 - 1)!\ (n_2 - 1)!}{(r_1 - 1)!\ (n_1 - r_1 - 1)!\ (n_1 + n_2 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j)!\ (n_1 + n_2 - r_1 - j - 2)!\ (j + 1 - r_2)}{(j + 1)!\ (n_2 - j - 1)!}$$

and the values of S_1 at the transition points follow.

(1) $S_1(r_1 + 1, n_1, r_2, n_2) - S_1(r_1, n_1, r_2, n_2)$

$$= \frac{K_1(n_1 - 1)!\ (n_2 - 1)!}{(r_1 - 1)!\ (n_1 - r_1 - 1)!\ (n_1 + n_2 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j)!\ (n_1 + n_2 - r_1 - j - 2)!\ (j + 1 - r_2)}{(j + 1)!\ (n_2 - 1 - j)!}$$

$$\left\{\frac{(r_1 + j + 1)(n_1 - r_1 - 1)}{r_1(n_1 + n_2 - r_1 - j - 2)} - 1\right\}$$

$$= \frac{K_1(n_1 - 1)!\ (n_2 - 1)!}{(r_1 - 1)!\ (n_1 - r_1 - 1)!\ (n_1 + n_2 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j)!\ (n_1 + n_2 - r_1 - j - 2)!\ (j + 1 - r_2)}{(j + 1)!\ (n_2 - 1 - j)!}$$

$$\left\{\frac{jn_1 - j + n_1 - 1 - r_1n_2}{r_1(n_1 + n_2 - r_1 - j - 2)}\right\} \tag{3.3.1}$$

Little may be said about the behaviour of the differences since the resulting summations of factorials are complex, however some idea may be obtained by considering the last term in brackets. In general, the difference is positive if the last term in brackets is positive for all values of j and negative if the last term is negative for all j. However, we are uncertain about the difference if the last term is positive for some values of j and negative for some others. If the inequality

$$r_1 \leqslant \frac{(j + 1)(n_1 - 1)}{n_2} \tag{3.3.2}$$

holds for all j then the difference (3.3.1) will always be positive. At $j = r_2$, RHS = $\dfrac{(r_2 + 1)(n_1 - 1)}{n_2}$; hence for points satisfying $r_1 \leqslant \dfrac{(r_2 + 1)(n_1 - 1)}{n_2}$ the difference (3.3.1) will be positive.

By the same argument the difference will be negative for points satisfying $r_1 > (n_1 - 1)$ which never hold.

A negative difference will obviously occur if $r_1 > \dfrac{(j + 1)(n_1 - 1)}{n_2}$ which violates the inequality $r_1 \leqslant n_1 - 1$.

The inequality will not necessarily define any points in the four dimensional integer space but consider the 'indifference' point (r_1, n_1, r_1, n_1) then the RHS of (3.3.2) with $j = r_1$ becomes

$r_1 + \left[\dfrac{n_1 - r_1 - 1}{n_1}\right]$ and the inequality holds.

(2) $S_1(r_1, n_1 + 1, r_2, n_2) - S_1(r_1, n_1, r_2, n_2)$

$$= \frac{K_1(n_1 - 1)!\,(n_2 - 1)!}{(r_1 - 1)!\,(n_1 - r_1 - 1)!\,(n_1 + n_2 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j)!\,(n_1 + n_2 - j - r_1 - 2)!\,(j + 1 - r_2)}{(j + 1)!\,(n_2 - j - 1)!}$$

$$\left\{\frac{n_1(n_1 + n_2 - j - r_1 - 1)}{(n_1 - r_1)(n_1 + n_2)} - 1\right\}$$

$$= \frac{K_1(n_1 - 1)!\,(n_2 - 1)!}{(r_1 - 1)!\,(n_1 - r_1 - 1)!\,(n_1 + n_2 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j)!\,(n_1 + n_2 - j - r_1 - 2)!\,(j + 1 - r_2)}{(j + 1)!\,(n_2 - j - 1)!}$$

$$\left\{\frac{-n_1 j - n_1 + r_1 n_2}{(n_1 - r_1)(n_1 + n_2)}\right\} \qquad (3.3.3)$$

Following the same argument as in (1), we investigate the last term in brackets of (3.3.3). The difference (3.3.3) will be positive if the inequality

$$r_1 \geqslant \frac{n_1(j + 1)}{n_2} \qquad (3.3.4)$$

holds for all j, which never occur since it violates the condition $r_1 \leqslant n_1 - 1$. However, the inequality

$$r_1 < \frac{n_1(j + 1)}{n_2}$$

will hold for all $j = r_2, \ldots, n_2 - 1$; if $r_1 < \dfrac{n_1(r_2 + 1)}{n_2}$,

giving negative values for the difference of risks.

(3) $S_1(r_1 + 1, n_1 + 1, r_2, n_2) - S_1(r_1, n_1, r_2, n_2)$

$$= \frac{K_1(n_1 - 1)!\,(n_2 - 1)!}{(r_1 - 1)!\,(n_1 - r_1 - 1)!\,(n_1 + n_2 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j)!\,(n_1 + n_2 - j - r_1 - 2)!\,(j + 1 - r_2)}{(j + 1)!\,(n_2 - j - 1)!}$$

$$\left\{\frac{n_1(r_1 + j + 1)}{r_1(n_1 + n_2)} - 1\right\}$$

$$= \frac{K_1(n_1 - 1)!\,(n_2 - 1)!}{(r_1 - 1)!\,(n_1 - r_1 - 1)!\,(n_1 + n_2 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j)!\,(n_1 + n_2 - j - r_1 - 2)!\,(j + 1 - r_2)}{(j + 1)!\,(n_2 - j - 1)!}$$

$$\left\{\frac{n_1(j + 1) - r_1 n_2}{r_1(n_1 + n_2)}\right\} \qquad (3.3.5)$$

Using the same agrument as for the last result, the difference (3.3.5) will be certainly positive if

$$r_1 < \frac{n_1(r_2 + 1)}{n_2}.$$

(4) $S_1(r_1, n_1, r_2 + 1, n_2) - S_1(r_1, n_1, r_2, n_2)$

$$= \frac{K_1(n_1 - 1)!\,(n_2 - 1)!}{(r_1 - 1)!\,(n_1 - r_1 - 1)!\,(n_1 + n_2 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j)!\,(n_1 + n_2 - j - r_1 - 2)!}{(j + 1)!\,(n_2 - j - 1)!}$$

$$\{(j - r_2) - (j + 1 - r_2)\} \qquad (3.3.7)$$

then the difference is negative for all points.

(5) $S_1(r_1, n_1, r_2, n_2 + 1) - S_1(r_1, n_1, r_2, n_2)$

$$= H_1 + \frac{K_1(n_1 - 1)!\,(n_2 - 1)!}{(r_1 - 1)!\,(n_1 - r_1 - 1)!\,(n_1 + n_2 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j)!\,(n_1 + n_2 - r_1 - j - 2)!\,(j + 1 - r_2)}{(j + 1)!\,(n_2 - j - 1)!}$$

$$\left\{\frac{n_2(n_1 + n_2 - r_1 - j - 1)}{(n_1 + n_2)(n_2 - j)} - 1\right\}$$

$$= H_1 + \frac{K_1(n_1 - 1)!\,(n_2 - 1)!}{(r_1 - 1)!\,(n_1 - r_1 - 1)!\,(n_1 + n_2 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j)!\,(n_1 + n_2 - r_1 - j - 2)!\,(j + 1 - r_2)}{(j + 1)!\,(n_2 - j - 1)!}$$

$$\left\{\frac{n_2(n_1 + n_2 - r_1 - j - 1) - (n_1 + n_2)(n_2 - j)}{(n_1 + n_2)(n_2 - j)}\right\} \qquad (3.3.8)$$

where

$$H_1 = \frac{K_1(n_1 - 1)!\ (r_1 + n_2)!\ (n_2 + 1 - r_2)}{(r_1 - 1)!\ (n_1 + n_2)!\ (n_2 + 1)}$$

If $r_1 \leqslant \dfrac{n_1 r_2 - n_2}{n_2}$ then the difference is positive and if

$r_1 > \dfrac{n_1(n_2 - 1) - n_2}{n_2}$ then the difference is negative. At the indifference point (r_1, n_1, r_1, n_1) the first inequality does not hold and the second holds for some r_1.

(6) $S_1(r_1, n_1, r_2 + 1, n_2 + 1) - S_1(r_1, n_1, r_2, n_2)$

$$= \frac{K_1(n_1 - 1)!\ n_2!}{(r_1 - 1)!\ (n_1 - r_1 - 1)!\ (n_1 + n_2)!} \times$$

$$\sum_{j=r_2+1}^{n_2} \frac{(r_1 + j)!\ (n_1 + n_2 - r_1 - j - 1)!\ (j - r_2)}{(j + 1)!\ (n_2 - j)!}$$

$$- S_1(r_1, n_1, r_2, n_2)$$

$$= \frac{K_1(n_1 - 1)!\ n_2!}{(r_1 - 1)!\ (n_1 - r_1 - 1)!\ (n_1 + n_2)!} \times$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j + 1)!\ (n_1 + n_2 - r_1 - j - 2)!\ (j + 1 - r_2)}{(j + 2)!\ (n_2 - j - 1)!}$$

$$- S_1(r_1, n_1, r_2, n_2)$$

$$= \frac{K_1(n_1 - 1)!\,(n_2 - 1)!}{(r_1 - 1)!\,(n_1 - r_1 - 1)!\,(n_1 + n_2 - 1)!} \times$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j)!\,(n_1 + n_2 - r_1 - j - 2)!\,(j + 1 - r_2)}{(j + 1)!\,(n_2 - j - 1)!} \times$$

$$\left\{\frac{n_2(r_1 + j + 1) - (n_1 + n_2)(j + 2)}{(n_1 + n_2)(j + 2)}\right\}. \qquad (3.3.10)$$

Then

$$S_1(r_1, n_1, r_2 + 1, n_2 + 1) - S_1(r_1, n_1, r_2, n_2) \geq 0$$

if

$$n_2(r_1 + j + 1) - (n_1 + n_2)(j + 2) \geq 0$$

or

$$r_1 \geq \frac{n_2 + n_1(j + 2)}{n_2} \qquad \text{for all } j = r_2, \ldots, n_2 - 1.$$

Further comments may be made here.

S_1 = stopping risk in making decision D_1: $p_1 \leq p_2$.

So behaviour is as expected; if r_1 increases, $p_1 \leq p_2$ becomes more unlikely. Similarly the behaviour of $S_1 - S_2$ has the expected properties as r_1, r_2 changes.

From numerical work we have confirmed that the following properties hold.

(1) S_1 increases as r_1 increases, r_2 is fixed.

(2) S_1 decreases as r_2 increases, r_1 is fixed.

(3) S_1 decreases as n_1 increases for fixed values of r_1 and r_2.

$$\text{i.e. } \lim_{n_1 \to \infty} S_1 = 0.$$

(4) S_1 for small r_1 decreases faster as r_2 increases than for large r_1.

(5) $S_1 - S_2$ increases at a constant rate as r_2 increases, for fixed r_1.

(6) $S_1 - S_2$ decreases at a constant rate as r_1 increases, for fixed r_2.

3.4 The fully Bayesian sequential scheme OPT_1

The fully Bayesian sequential scheme, denoted OPT_1, to select the better of two Binomial populations is presented. The Bayesian decision-theoretic formulation given in section 3.1 in conjunction with the dynamic programming technique is used to construct this procedure. The observations are taken from the populations sequentially one at a time.

At each point (r_1, n_1, r_2, n_2) in four dimensional integer space, the optimal decision to stop or continue is made by comparing the posterior expected loss with expected risk of taking one more observation.

At the point (r_1, n_1, r_2, n_2), let

$S_1(r_1, n_1, r_2, n_2)$ be the stopping risk of making the terminal decision D_1,

$S_2(r_1, n_1, r_2, n_2)$ be the stopping risk of making the terminal decision D_2,

$B_1(r_1, n_1, r_2, n_2)$ be the risk of taking one further observation from π_1 and proceeding optimally thereafter, termed the continuation risk for π_1,

$B_2(r_1, n_1, r_2, n_2)$ be the risk of taking one further observation from π_2 and proceeding optimally thereafter, termed the continuation risk for π_2,

$D(r_1, n_1, r_2, n_2)$ be the minimum risk (giving the optimal policy).

We assume that the cost of sampling each observation from population π_i is constant and denoted by C_i $(i = 1, 2)$.

At each point, there are four possible transitions

$(r_1 + 1, n_1 + 1, r_2, n_2)$ with probability $\frac{r_1}{n_1}$,

$(r_1, n_1 + 1, r_2, n_2)$ with probability $\frac{n_1 - r_1}{n_1}$ if the next observation is taken from π_1 and $(r_1, n_1, r_2 + 1, n_2 + 1)$ with probability $\frac{r_2}{n_2}$, $(r_1, n_1, r_2, n_2 + 1)$ with probability $\frac{n_2 - r_2}{n_2}$ if the next observation is taken from π_2, then

$$B_1(r_1, n_1, r_2, n_2) = C_1 + \frac{r_1}{n_1} D(r_1 + 1, n_1 + 1, r_2, n_2) + \frac{n_1 - r_1}{n_1} D(r_1, n_1 + 1, r_2, n_2) \qquad (3.4.1)$$

and

$$B_2(r_1, n_1, r_2, n_2) = C_2 + \frac{r_2}{n_2} D(r_1, n_1, r_2 + 1, n_2 + 1)$$

$$+ \frac{n_2 - r_2}{n_2} D(r_1, n_1, r_2, n_2 + 1). \tag{3.4.2}$$

Knowing $S_1(r_1, n_1, r_2, n_2)$, $S_2(r_1, n_1, r_2, n_2)$, $B_1(r_1, n_1, r_2, n_2)$ and $B_2(r_1, n_1, r_2, n_2)$, the equation for $D(r_1, n_1, r_2, n_2)$ is given by

$$D(r_1, n_1, r_2, n_2) = \min(S(r_1, n_1, r_2, n_2), B(r_1, n_1, r_2, n_2)) \tag{3.4.3}$$

where,

$$S(r_1, n_1, r_2, n_2) = \min(S_1(r_1, n_1, r_2, n_2), S_2(r_1, n_1, r_2, n_2))$$

$$B(r_1, n_1, r_2, n_2) = \min(B_1(r_1, n_1, r_2, n_2), B_2(r_1, n_1, r_2, n_2)).$$

Suppose that the procedure is truncated at N observations, (that is the maximum number of observations that can be taken through the whole sampling procedure is N), then the dynamic programming equations above are used successively from this end point to the origin to partition the four dimensional integer space into stopping and continuation points. Due to the dynamic programming technique of computation it is not known which points are reachable by any simple path starting at (a_1, b_1, a_2, b_2) until this origin is reached.

The effect of truncation is to increase B_1, B_2 in general and hence produce more stopping points than an optimal procedure if it exists, the fully optimal procedure being that where a maximum sample size exists such that all possible points in four dimensional space for this sample size are stopping points. A further discussion of the effect of truncation is given in section (3.7).

Similar equations can be found in Lindley and Barnett (1965), Freeman (1972), and Jones (1974).

The form of the stopping risks precludes any analytical methods of determining whether an optimal maximum sample size exists beyond which no reduction of the overall risk is possible. Computational experience however suggests that such exists. If it can be shown that

$$\lim_{N\to\infty} \min[S_1, S_2] = 0 \text{ for all } r_1, r_2 \text{ and } (n_1 + n_2) - (b_1 + b_2) = N$$

then it follows that

$$\lim_{N\to\infty} [B_1, B_2] = C_1 \text{ or } C_2,$$

since $D \leqslant \min[S_1, S_2]$ and hence an optimal maximum sample size exists. The appearance of N! in the stopping risks suggests that such a limit may exist.

The stopping rule of the optimal scheme can be described as follows.

At the point (r_1, n_1, r_2, n_2),

(i) Stop sampling and make that terminal decision with smaller risk as soon as

$$D(r_1, n_1, r_2, n_2) = S(r_1, n_1, r_2, n_2) \leqslant B(r_1, n_1, r_2, n_2).$$

(ii) If no terminal decision has been reached before N, then terminate sampling and take that terminal decision with smaller risk.

(iii) If

$$D(r_1, n_1, r_2, n_2) = B(r_1, n_1, r_2, n_2) < S(r_1, n_1, r_2, n_2)$$

then continue sampling with the population which has smaller continuation risk.

The terminal decision is as follows.

At the point (r_1, n_1, r_2, n_2),

we choose decision D_1 and select population π_1 as better if

$$S_1(r_1, n_1, r_2, n_2) \leq S_2(r_1, n_1, r_2, n_2)$$

and we choose decision D_2 and select population π_2 as better if

$$S_1(r_1, n_1, r_2, n_2) > S_2(r_1, n_1, r_2, n_2).$$

3.5 Some properties of OPT_1

In this section we give some properties of OPT_1. These properties hold for all loss functions.

Result 1

If at the transition points $(r_1 + 1, n_1 + 1, r_2, n_2)$ and $(r_1, n_1 + 1, r_2, n_2)$, $D = S_1$ (or both are S_1 points) then the point (r_1, n_1, r_2, n_2) cannot be a continue 1 (B_1) point.

To show this is true, first consider S_1 under the linear loss function. The form (3.2.4) of S_1 can be rewritten as

$$S_1(r_1, n_1, r_2, n_2) = \frac{K_1(n_1 - 1)!\,(n_2 - 1)!}{(r_1 - 1)!\,(n_1 - r_1 - 1)!\,(n_1 + n_2 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j)!\,(n_1 + n_2 - j - r_1 - 2)!\,(j + 1 - r_2)}{(j + 1)!\,(n_2 - j - 1)!}. \tag{3.5.1}$$

Then,

$$B_1(r_1, n_1, r_2, n_2) = C_1 + \frac{r_1}{n_1} S_1(r_1 + 1, n_1 + 1, r_2, n_2)$$

$$+ \frac{n_1 - r_1}{n_1} S_1(r_1, n_1 + 1, r_2, n_2) \tag{3.5.2}$$

$$= C_1 + \frac{K_1(n_1 - 1)!\,(n_2 - 1)!}{(r_1 - 1)!\,(n_1 - r_1 - 1)!\,(n_1 + n_2)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j + 1)!\,(n_1 + n_2 - j - r_1 - 2)!\,(j + 1 - r_2)}{(j + 1)!\,(n_2 - j - 1)!}$$

$$+ \frac{K_1(n_1 - 1)!\,(n_2 - 1)!}{(r_1 - 1)!\,(n_1 - r_1 - 1)!\,(n_1 + n_2)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j)!\,(n_1 + n_2 - j - r_1 - 1)!\,(j + 1 - r_2)}{(j + 1)!\,(n_2 - j - 1)!}$$

$$= C_1 + \frac{K_1(n_1 - 1)!\,(n_2 - 1)!}{(r_1 - 1)!\,(n_1 - r_1 - 1)!\,(n_1 + n_2 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j)!\ (n_1 + n_2 - j - r_1 - 2)!\ (j + 1 - r_2)}{(j + 1)!\ (n_2 - j - 1)!}$$

$$\left\{\frac{(r_1 + 1 + j) + n_1 + n_2 - j - r_1 - 1}{(n_1 + n_2)}\right\}$$

$$= C_1 + S_1(r_1, n_1, r_2, n_2). \qquad (3.5.3)$$

Under the quadratic loss function,

$$B_1(r_1, n_1, r_2, n_2) = C_1 + \frac{K_1(n_1 - 1)!\ (n_2 - 1)!}{(r_1 - 1)!\ (n_1 - r_1 - 1)!\ (n_1 + n_2)!}$$

$$\sum_{j=r_2}^{n_2-1} \left\{\frac{(r_1 + j + 1)!\ (n_1 + n_2 - j - r_1 - 2)!}{(j + 2)!\ (n_2 - j - 1)!}\right.$$

$$\left. \times\ (j^2 + 3j - 2jr_2 + r_2^2 - 3r_2 + 2)\right\}$$

$$(3.5.4)$$

and under the constant loss function,

$$B_1(r_1, n_1, r_2, n_2) = C_1 + \frac{K_1(n_1 - 1)!\ (n_2 - 1)!}{(r_1 - 1)!\ (n_1 - r_1 - 1)!\ (n_1 + n_2 - 2)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{(r_1 + j - 1)!\ (n_1 + n_2 - r_1 - j - 2)!}{j!\ (n_2 - 1 - j)!}.$$

$$(3.5.5)$$

<u>Result 2</u>

If at the transition points $(r_1, n_1, r_2 + 1, n_2 + 1)$ and $(r_1, n_1, r_2, n_2 + 1)$, $D = S_2$ (or both S_2 points) then at the point (r_1, n_1, r_2, n_2),

$$B_2(r_1, n_1, r_2, n_2) = C_2 + S_2(r_1, n_1, r_2 n_2) \tag{3.5.6}$$

(it follows from Result 1).

<u>Result 3</u>

If the transition points $(r_1 + 1, n_1 + 1, r_2, n_2)$ is an S_1 point and $(r_1, n_1 + 1, r_2, n_2)$ is an S_2 point then the point (r_1, n_1, r_2, n_2) is not an S_1 point if

$$\left[\frac{n_1 - r_1}{n_1}\right][S_1(r_1, n_1 + 1, r_2, n_2) - S_2(r_1, n_1 + 1, r_2, n_2)] > C_1 \tag{3.5.7}$$

proof.

From equation (3.4.1), we get

$$\begin{aligned} B_1(r_1, n_1, r_2, n_2) &= C_1 + \frac{r_1}{n_1} S_1(r_1 + 1, n_1 + 1, r_2, n_2) \\ &\quad + \left(\frac{n_1 - r_1}{n_1}\right) S_2(r_1, n_1 + 1, r_2, n_2) \\ &= C_1 + \frac{r_1}{n_1} S_1(r_1 + 1, n_1 + 1, r_2, n_2) \\ &\quad + \left(\frac{n_1 - r_1}{n_1}\right) S_1(r_1, n_1 + 1, r_2, n_2) \\ &\quad - \left(\frac{n_1 - r_1}{n_1}\right) \{S_1(r_1, n_1 + 1, r_2, n_2) \\ &\quad - S_2(r_1, n_1 + 1, r_2, n_2)\}. \end{aligned} \tag{3.5.8}$$

From Result 1, (3.5.8) would become

$$B_1(r_1, n_1, r_2, n_2) = C_1 + S_1(r_1, n_1, r_2, n_2)$$

$$- \left(\frac{n_1 - r_1}{n_1}\right) \{S_1(r_1, n_1 + 1, r_2, n_2)$$

$$- S_2(r_1, n_1 + 1, r_2, n_2)\}. \qquad (3.5.9)$$

Since the point (r_1, n_1, r_2, n_2) is not S_1 point if

$$B_1(r_1, n_1, r_2, n_2) - S_1(r_1, n_1, r_2, n_2) < 0, \qquad (3.5.10)$$

hence (3.5.7) follows.

Result 4

If the transition point $(r_1, n_1, r_2 + 1, n_2 + 1)$ is an S_1 point and $(r_1, n_1, r_2, n_2 + 1)$ is an S_2 point then the point (r_1, n_1, r_2, n_2) is not an S_2 point if

$$\frac{r_2}{n_2} [S_2(r_1, n_1, r_2 + 1, n_2 + 1) - S_1(r_1, n_1, r_2 + 1, n_2 + 1)] > C_2$$

$$(3.5.11).$$

follows from Result 3.

3.6 The influence of loss constants, sampling costs and prior information on the optimal overall risk

As the optimal overall risk is a function of the sample size N, the loss constants K_1 and K_2, the sampling costs C_1 and C_2 and the prior parameters (a_1, b_1) and (a_2, b_2), therefore it varies as they vary. In this section we present some numerical work to investigate the effect of these factors on the optimal overall risk of the optimal scheme OPT_1 for constant, linear and quadratic loss functions. For ease of notation we denote the optimal risk at the origin $D(a_1, b_1, a_2, b_2)$ by $D(N)$.

The numerical results are given in Tables (3.1 - 3.3). The relation between $\frac{D(N)}{D(10)}$ and N for various loss functions is graphed in Figure (3.1). Graphical illustrations of the effect of prior information on p_1 and p_2, the loss constants and the sampling costs on $\frac{D(N)}{D(10)}$ for different values of N using linear loss function are given in Figures (3.2 - 3.4).

From Tables (3.1 - 3.4), we note that as N increases the optimal overall risk D(N) decreases, but the rate of decrease of D(N) becomes smaller. In terms of a graph, as shown in Figures (3.1 - 3.4), the $\frac{D(N)}{D(10)}$ curve is beginning to decrease. The rate of decrease reflects the fact that as the sample size gets larger and larger, the additional value of each observation becomes smaller and smaller. We note that the rate of decrease of D(N) is rather quicker in the case of quadratic losses, then linear losses come next. The behaviour of D(N) as a function of N is important in determining the optimal sample size for the scheme and will be discussed in section 3.7.

Table (3.1) shows, for N = 10(10)100 and for constant, linear and quadratic loss functions, how the prior probabilities on p_1, p_2 affect D(N). It is clear from this table that D(N) becomes smaller with stronger prior probabilities. For instance, if we have $C_1 = C_2 = 1$, $K_1 = K_2 = 1000$ with quadratic loss function and N = 20, then

Table (3.1)

The influence of the prior information on the optimal overall risk in OPT_1, under different loss functions and different values of N, when

$$K_1 = K_2 = 1000, \qquad C_1 = C_2 = 1.$$

Prior Probs.	N	Loss Function		
		Constant	Linear	Quadratic
Be(1, 2) v Be(1, 2)	10	183.762	32.9121	13.0008
	20	143.877	26.1292	12.2307
	30	126.675	24.7752	12.2118
	40	117.008	24.4258	
	50	110.905	24.3341	
	60	106.732	24.3125	
	70	103.771	24.3081	
	80	101.572	24.3074	
	90	99.8998	24.3073	
	100	98.6122		
Be(1, 3) v Be(1, 2)	10	174.858	30.0599	11.6637
	20	140.086	24.8118	11.2205
	30	124.206	23.6925	11.2118
	40	115.127	23.4026	
	50	109.319	23.3283	
	60	106.329	23.3112	
	70	102.479	23.3079	
	80	100.352	23.3073	
	90	98.7325		
	100	97.4836		

Table (3.1) (continued)

The influence of the prior information on the optimal overall risk in OPT_1, for different loss functions and different values of N, when

$$K_1 = K_2 = 1000, \qquad C_1 = C_2 = 1.$$

Prior Probs.	N	Loss Function		
		Constant	Linear	Quadratic
Be(1, 5) v Be(1, 2)	10	140.015	21.0814	7.40078
	20	116.084	18.5123	7.29749
	30	104.349	17.9854	7.29626
	40	97.4394	17.8609	
	50	92.9423	17.8318	
	60	89.8292	17.8256	
	70	87.5744	17.8245	
	80	85.8810	17.8244	
	90	84.5901		
	100	83.5910		

D(20) = 12.307 for uniform priors (1, 2) v (1, 2)

and

D(20) = 11.2205 for Beta priors (1, 3) v (1, 2).

Table (3.2) demonstrates how the loss constants K_1, K_2 have a considerable effect on the values of D(N).

Note: The missing entries in each column in the tables have the same value of the last entry in that column.

Table (3.2)

The influence of the loss constants K_1, K_2 on the optimal overall risk in OPT_1 for different loss functions and different values of N when $C_1 = C_2 = 1$ with uniform priors.

K_1, K_2	N	Loss Function		
		Constant	Linear	Quadratic
$K_1 = 1000$	10	183.762	32.9121	13.0008
$K_2 = 1000$	20	143.877	26.1292	12.2307
	30	126.675	24.7752	12.2118
	40	117.008	24.4258	
	50	110.905	24.3341	
	60	106.732	24.3125	
	70	103.771	24.3081	
	80	101.572	24.3074	
	90	99.8998	24.3073	
	100	98.6122		
$K_1 = 10000$	10	1771.6500	265.7290	72.3140
$K_2 = 10000$	20	1317.7600	157.1240	40.4851
	30	1105.1900	118.6620	32.9966
	40	974.660	99.6494	30.5766
	50	884.890	88.8198	29.6940
	60	818.346	82.0480	29.3611
	70	766.669	77.5853	29.2399
	80	725.105	74.5057	29.1987
	90	690.847	72.3277	29.1856
	100	661.992	70.7581	29.1767

Table (3.2) (continued)

The influence of the loss constants K_1, K_2 on the optimal overall risk in OPT_1 for different loss functions and different values of N when $C_1 = C_2 = 1$ with uniform priors.

K_1, K_2	N	Loss Function		
		Constant	Linear	Quadratic
$K_1 = 10000$	10	366.324	69.3102	23.3827
$K_2 = 1000$	20	293.164	48.9628	18.3379
	30	254.508	41.9887	17.6687
	40	231.011	39.2553	17.5647
	50	215.130	37.9208	17.5564
	60	203.536	37.2767	
	70	194.695	36.9618	
	80	187.874	36.7994	
	90	182.268	36.7181	
	100	177.764	36.6783	

Table (3.3)

The influence of the sampling costs on the optimal overall risk in OPT_1, for different loss functions and different values of N when

$K_1 = K_2 = 1000$ with uniform priors.

C_1, C_2	N	Loss Function: Constant	Linear	Quadratic
$C_1 = 1$	10	183.762	32.9121	13.0008
$C_2 = 1$	20	143.877	26.1292	12.2307
	30	126.675	24.7752	12.2118
	40	117.008	24.4258	
	50	110.905	24.3341	
	60	106.732	24.3125	
	70	103.771	24.3081	
	80	101.572	24.3074	
	90	99.8998	24.3073	
	100	98.6122		
$C_1 = 5$	10	192.258	40.2824	17.8296
$C_2 = 1$	20	161.287	37.3067	17.7039
	30	149.751	37.0584	
	40	144.145	37.0402	
	50	141.130	37.0393	
	60	139.376		
	70	138.309		
	80	137.641		
	90	137.219		
	100	136.949		

Table (3.3) (continued)

The influence of the sampling costs on the optimal overall risk in OPT_1, for different loss functions and different values of N when

$K_1 = K_2 = 1000$ with uniform priors.

C_1, C_2	N	Loss Function: Constant	Linear	Quadratic
$C_1 = 5$	10	211.257	55.9591	29.5774
$C_2 = 5$	20	187.764	55.4207	
	30	181.225	55.4180	
	40	179.083		
	50	178.213		
	60	177.876		
	70	177.749		
	80	177.701		
	90	177.685		
	100	177.679		

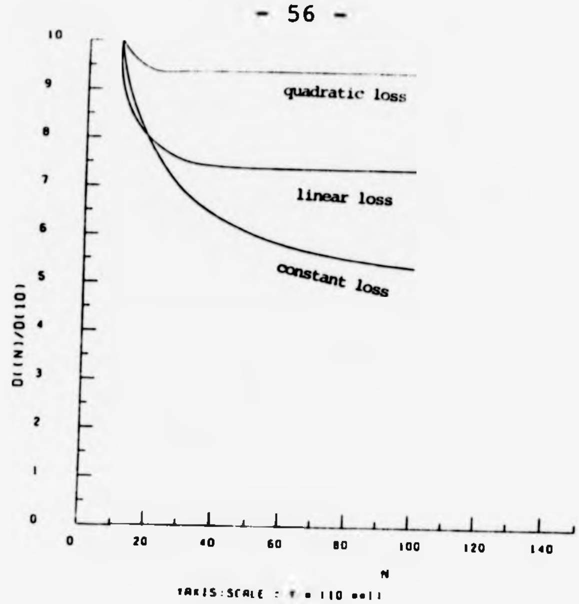

Fig. 3.1 D(N)/D(10) as a function of N for OPT_1 under different loss functions where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$ with uniform priors.

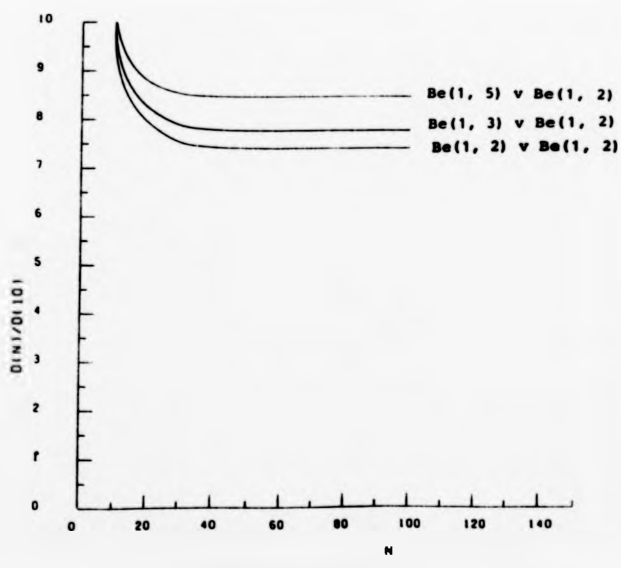

Fig. 3.2 D(N)/D(10) as a function of N for OPT_1 under different priors where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$ with linear loss function.

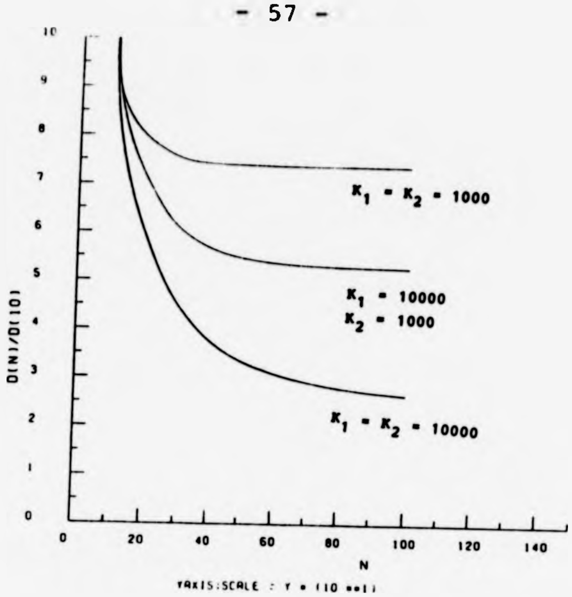

Fig. 3.3 D(N)/D(10) as a function of N for OPT_1 under different loss constants where $C_1 = C_2 = 1$ with uniform priors and linear loss function.

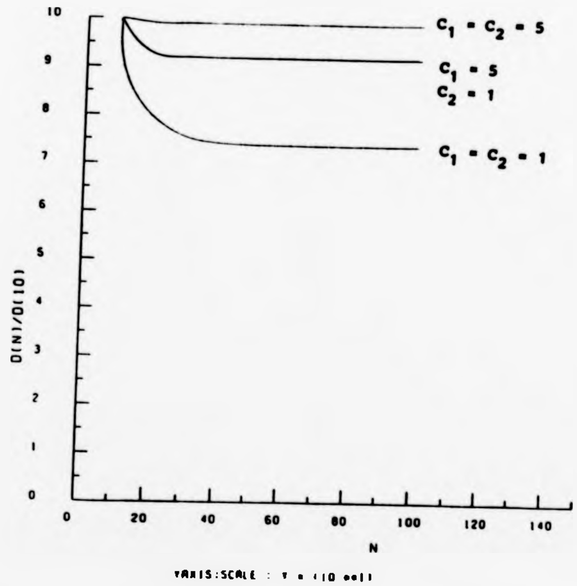

Fig. 3.4 D(N)/D(10) as a function of N for OPT_1 under different sampling costs where $K_1 = K_2 = 1000$ with uniform priors and linear loss function.

3.7 Determination of the optimal maximum sample size

Suppose that the purely sequential scheme is truncated at a maximum of N observations. If we wish to determine the optimal maximum sample size, denoted by M*, we compute the optimal overall risks of sampling, D(N), for various values of N. Then,

$$D(M^*) \leqslant D(N) \qquad \text{for all } N.$$

This will be the first value of N such that all points are stopping points giving a constant value of D(N) for $N \geqslant M^*$.

In Table (3.7), where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, uniform priors (1, 2) v (1, 2) and linear loss function, $M^* = 81$.

The percentage reduction in D(N), denoted by (DR%), gained by using the optimal maximum sample size M* is given by

$$(DR\%) = \frac{D(N) - D(M^*)}{D(M^*)} \times 100.$$

The values of M* and (DR%) are given in Tables (3.7) and (3.9) ((3.10) and (3.12)) under the linear (quadratic) loss function. We note that (DR%) is not very sensitive to changes in N for large N.

Sometimes there is no optimal maximum sample size over the range of values of N considered, then we can use the percentage decrease in risk over the previous sample size, denoted by (DD%) and given by

$$(DD\%) = \frac{D(N) - D(N - 1)}{D(N)} \times 100,$$

to show that D(N) is decreasing at a slower rate. For instance, in Tables (3.5), (3.8) and (3.11), where N = 20, $K_1 = K_2 = 10000$, $C_1 = C_2 = 1$ with uniform priors, the values of (DD%) are 25.6%, 40.87% and 44.01% under constant, linear and quadratic loss functions respectively.

When the prior distributions are different from uniform priors then provided these new priors are proper and have integer parameters the optimal maximum sample size is naturally reduced. Some numerical examples, given in Tables (3.7) and (3.10) under linear and quadratic loss functions respectively, display this behaviour. Table (3.4) shows that M* > 100 under both priors Be(1, 3) v Be(1, 2) and Be(1, 2) v Be(1, 2); however the rate of decrease in (DD%) is smaller under the first set of priors and that provides numerical evidence that the value of M* using Be(1, 3) v Be(1, 2) is less than M* using Be(1, 2) v Be(1, 2).

Increasing the value of loss constants K_1 and K_2 naturally increases the optimal maximum sample size. In Tables (3.5), (3.8) and (3.11), where constant, linear and quadratic loss functions are used respectively, the behaviour of (DD%) shows that M* under $K_1 = K_2 = 1000$ is less than M* under $K_1 = K_2 = 10000$. In contrast, if observations cost more then M* will decrease. The results of Table (3.9) and (3.12), produced under linear and quadratic loss functions respectively, numerically support this fact. In addition, in Table (3.6) where constant loss function is used, the decrease in (DD%) is faster under the case $C_1 = C_2 = 5$ than the case $C_1 = C_2 = 1$ which again illustrates that M* is less under the first case.

It is possible for M* to be zero, in which case a decision should be made without further sampling. For example, if

$K_1 = 1000$, $K_2 = 10000$ with Beta priors (1, 10) v (1, 2) then $D(N) = 9.0901$ for all N and hence $M^* = 0$.

As noted in Lindley and Barnett (1965) and in considering the above values of (DD%), truncation produces very little increase in risk which suggests it is worth using in many cases.

Table (3.4)

D(N) and the percentage decrease in risk over the previous sample size (DD%) in OPT_1, for $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, constant loss function with two sets of priors.

N	Be(1, 2) v Be(1, 2)		Be(1, 3) v Be(1, 2)	
	D(N)	(DD%)	D(N)	(DD%)
10	183.762		174.858	
20	143.877	21.7	140.086	19.89
30	126.675	11.96	124.206	11.34
40	117.008	7.63	115.127	7.31
50	110.905	5.2	109.319	5.04
60	106.732	3.76	106.329	2.74
70	103.771	2.77	103.479	2.68
80	101.572	2.12	100.352	2.08
90	99.899	1.65	98.732	1.61
100	98.612	1.29	97.483	1.26

Table (3.5)

D(N) and (DD%) in OPT_1 for $C_1 = C_2 = 1$, uniform priors with constant loss function and two sets of K_1, K_2 values.

N	$K_1 = K_2 = 1000$		$K_1 = K_2 = 10000$	
	D(N)	(DD%)	D(N)	(DD%)
10	183.762		1771.650	
20	143.877	21.7	1317.760	25.62
30	126.675	11.96	1105.190	16.13
40	117.008	7.63	974.660	11.81
50	110.905	5.2	884.890	9.21
60	106.732	3.76	818.346	7.5
70	103.771	2.77	766.669	6.31
80	101.572	2.12	725.105	5.42
90	99.899	1.65	690.847	4.72
100	98.612	1.29	661.992	4.18

Table (3.6)

D(N) and (DD%) in OPT_1 for $K_1 = K_2 = 1000$, uniform priors with constant loss function and two sets of C_1, C_2 values.

N	$C_1 = C_2 = 1$		$C_1 = C_2 = 5$	
	D(N)	(DD%)	D(N)	(DD%)
10	183.762		211.257	
20	143.877	21.7	187.764	11.12
30	126.675	11.96	181.225	3.48
40	117.008	7.63	179.083	1.18
50	110.905	5.2	178.213	.486
60	106.732	3.76	177.876	.189
70	103.771	2.77	177.749	.070
80	101.572	2.12	177.701	.027
90	99.899	1.65	177.685	.009
100	98.612	1.29	177.679	.003

Table (3.7)

D(N) and (DR%) in OPT_1 for $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear loss function with two sets of priors.

N	Be(1, 2) v Be(1, 2)		N	Be(1, 3) v Be(1, 2)	
	D(N)	(DR%)		D(N)	(DR%)
10	32.0121	35.4000	10	30.0599	28.9720
20	26.1292	7.4953	20	24.8118	6.4550
30	24.7752	1.9250	30	23.6925	1.6527
40	24.4258	0.4875	40	23.4026	0.4089
50	24.3341	0.1102	50	23.3283	0.0910
60	24.3125	0.0213	60	23.3112	0.0167
70	24.3081	0.0032	70	32.3079	0.0025
80	24.3074	0.0004	76	23.3073	0.0000
81	24.3073	0.0000			

Table (3.8)

D(N) and (DD%) in OPT_1 for $C_1 = C_2 = 1$, uniform priors, $K_1 = K_2 = 10000$, under linear loss function.

N	D(N)	(DD%)
10	265.7290	
20	157.1240	40.87
30	118.6620	24.48
40	99.6494	15.02
50	88.8198	10.87
60	82.0480	7.62
70	77.5853	5.44
80	74.5057	3.97
90	72.3277	2.92
100	70.7581	2.17

Table (3.9)

D(N) and (DR%) in OPT_1 for $K_1 = K_2 = 1000$, uniform priors with linear loss function and two sets of C_1, C_2 values.

N	$C_1 = C_2 = 1$		N	$C_1 = C_2 = 5$	
	D(N)	(DR%)		D(N)	(DR%)
10	32.9121	35.4000	10	55.9591	.9748
20	26.1292	7.4953	20	55.4207	.0032
30	24.7752	1.9250	21	55.4189	0.000
40	24.4258	0.4875			
50	24.3341	0.1102			
60	24.3125	0.0213			
70	24.3081	0.0032			
80	24.3074	0.0004			
81	24.3073	0.000			

Table (3.10)

D(N) and (DR%) in OPT_1 for $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, quadratic loss function with two sets of priors.

N	Be(1, 2) v Be(1, 2)		N	Be(1, 3) v Be(1, 2)	
	D(N)	(DR%)		D(N)	(DR%)
10	13.0008	6.4522	10	11.6637	4.0213
20	12.2307	0.1466	20	11.2205	0.0686
25	12.2128	0.000	24	11.2128	0.000

Table (3.11)

D(N) and (DD%) in OPT_1 for $C_1 = C_2 = 1$, uniform priors, $K_1 = K_2 = 10000$, under quadratic loss function.

N	D(N)	(DD%)
10	72.3140	
20	40.4851	44.01
30	32.9966	18.50
40	30.5766	7.33
50	29.6940	2.89
60	29.3611	1.12
70	29.2399	.41
80	29.1987	.14
90	29.1856	.04
100	29.1767	.03

Table (3.12)

D(N) and (DR%) in OPT_1 for $K_1 = K_2 = 1000$, uniform priors with quadratic loss function and two sets of C_1, C_2 values.

	$C_1 = C_2 = 1$			$C_1 = C_2 = 5$	
N	D(N)	(DR%)	N	D(N)	(DR%)
10	13.0008	6.452	5	29.7156	0.467
20	12.2307	0.146	7	29.5774	0.000
25	12.2118	0.000			

CHAPTER 4

OPTIMAL (BAYESIAN) GROUP SEQUENTIAL AND OPTIMAL FIXED SAMPLE SIZE SCHEMES FOR SELECTING THE BETTER OF TWO BINOMIAL POPULATIONS

4.0 Summary

In this chapter optimal (Bayesian) group sequential and optimal fixed sample size schemes are developed. The construction and the description of the group sequential schemes with some numerical results are given in section 4.1. Section 4.2 discusses the optimal fixed sample size scheme. Section 4.3 contains some discussion and comparison in terms of relative efficiencies of both schemes with respect to the fully sequential scheme OPT_1.

4.1 Bayesian group sequential scheme OPT_2

4.1.1 Construction and description of the procedure

Suppose that observations are taken sequentially. The decision to stop or continue is made at a regular interval, where blocks of observations with equal-size 2n are taken, n on each population with a maximum overall sample size of N which is a multiple of 2n. This scheme is a practical formulation that retains most of the advantages of sequential analysis, particularly the economy in sample size, together with fixed sample sizes with its simplicity of use, saving time when applied in delayed response situations for example.

An important application of this scheme is in clinical trials with sequential patient arrival, where fixed sample size designs are unjustified on ethical grounds and sequential designs taking observations one at a time are often impracticable (Pocock (1977)). Therefore the main objective of this scheme is to reduce the number of patients on an inferior treatment by early termination of the sampling.

Let G be the number of groups such that

$$N = 2n * G. \tag{4.1.1}$$

Using the dynamic programming technique with the decision-theoretic formulation given in section 3.1 one can obtain the optimal decision at each point in the four dimensional integer space (r_1, n_1, r_2, n_2). A decision to stop or continue is made by comparing the posterior expected loss (stopping risk) with the expected risk of taking 2n more observations and proceeding optimality thereafter (continuation risk).

At the point (r_1, n_1, r_2, n_2), the stopping risks of taking decisions D_1: $p_1 \leq p_2$ and D_2: $p_1 > p_2$, denoted by S_1 and S_2 respectively, are found in Chapter 3. Furthermore, let

B_G^+ = continuation risk using both populations

D_G^+ = minimum risk attainable (giving the optimal policy)

C_1, C_2 = the constant costs of sampling an observation from populations π_1 and π_2 respectively.

The process can be represented in four dimensional space (r_1, n_1, r_2, n_2). There are $[(n + 1) \times (n + 1)]$ possible transitions $(r_1 + j, n_1 + n, r_2 + k, n_2 + n)$, with predictive probabilities

$$E\left[\binom{n}{j}\binom{n}{k}p_1^{\,j}\,q_1^{\,n-j}\,p_2^{\,k}\,q_2^{\,n-k}\right] \tag{4.1.2}$$

of reaching $(r_1 + j,\ n_1 + n,\ r_2 + k,\ n_2 + n)$ from $(r_1,\ n_1,\ r_2,\ n_2),\ 0 \leq j,\ k \leq n$.

Then the dynamic programming equations are

$$\begin{aligned} B_G^{+}(r_1, n_1, r_2, n_2) = n(C_1 + C_2) + \sum_{j=0}^{n}\sum_{k=0}^{n}\binom{n}{j}\binom{n}{k} \\ E\left[p_1^{\,j}\,q_1^{\,n-j}\,p_2^{\,k}\,q_2^{\,n-k}\right] \\ D_G^{+}(r_1 + j,\ n_1 + n,\ r_2 + k,\ n_2 + n) \end{aligned} \tag{4.1.3}$$

where

$$E\left[p_1^{\,j}\,q_1^{\,n-j}\,p_2^{\,k}\,q_2^{\,n-k}\right] = E\left[p_1^{\,j}(1 - p_1)^{n-j}\right]\,E\left[p_2^{\,k}(1 - p_2)^{n-k}\right],$$

(since p_1 and p_2 are independent)

and

$$\begin{aligned} E\left[p_1^{\,j}(1 - p_1)^{n-j}\right] &= \int_0^1 p_1^{\,j}(1 - p_1)^{n-j}\,\pi(p_1|r_1, n_1)\,dp_1 \\ &= \frac{(n_1 - 1)!}{(r_1 - 1)!\,(n_1 - r_1 - 1)!}\int_0^1 p_1^{\,r_1+j-1} \\ &\quad (1 - p_1)^{n_1+n-j-r_1-1}\,dp_1 \\ &= \frac{(n_1 - 1)!\,(r_1 + j - 1)!\,(n_1 + n - j - r_1 - 1)!}{(r_1 - 1)!\,(n_1 - r_1 - 1)!\,(n_1 + n - 1)!} \end{aligned}$$

Similarly,

$$E\left[p_2^{\,k}(1 - p_2)^{n-k}\right] = \frac{(n_2 - 1)!\,(r_2 + k - 1)!\,(n_2 + n - k - r_2 - 1)!}{(r_2 - 1)!\,(n_2 - r_2 - 1)!\,(n_2 + n - 1)!}$$

Giving S_1, S_2 and B_G^+ at the point (r_1, n_1, r_2, n_2), the minimum risk D_G^+ is given by

$$D_G^+(r_1, n_1, r_2, n_2) = \min[S_1(r_1, n_1, r_2, n_2), S_2(r_1, n_1, r_2, n_2),$$
$$B_G^+(r_1, n_1, r_2, n_2)]. \qquad (4.1.4)$$

The optimal overall risk can be obtained by working backwards from the maximum sample size N to origin, where $d_1 + d_2 = 0$, using the above dynamic programming equations. These equations give the stopping rule for this scheme as follows.

(i) Stop sampling and take that terminal decision with smaller risk as soon as

$$D_G^+(r_1, n_1, r_2, n_2) = \min(S_1(r_1, n_1, r_2, n_2),$$
$$S_2(r_1, n_1, r_2, n_2))$$
$$\leq B_G^+(r_1, n_1, r_2, n_2). \qquad (4.1.5)$$

(ii) Continue sampling with both populations if

$$D_G^+(r_1, n_1, r_2, n_2) = B_G^+(r_1, n_1, r_2, n_2)$$
$$< \min(S_1(r_1, n_1, r_2, n_2),$$
$$S_2(r_1, n_1, r_2, n_2)).$$

(iii) If no decision has been reached before N, then stop and take terminal decision as soon as N is reached.

The terminal decision rule:

At the point (r_1, n_1, r_2, n_2),

choose D_1 and select π_1 as better population if $S_1 \leq S_2$

choose D_2 and select π_2 as better population if $S_1 > S_2$.

Two extreme cases are given as follows.

1. The value n = 1 gives Wald-type sampling, where pairs of observations are taken at a time sequentially. In this case the continuation risk, denoted by B_2^+, is given by

$$\begin{aligned} B_2^+(r_1, n_1, r_2, n_2) = (C_1 + C_2) &+ \frac{r_1}{n_1}\left[\left(\frac{n_2 - r_2}{n_2}\right)\right. \\ & D_2^+(r_1 + 1, n_1 + 1, r_2, n_2 + 1) \\ & \left. + \frac{r_2}{n_2} D_2^+(r_1 + 1, n_1 + 1, r_2 + 1, n_2 + 1)\right] \\ & + \left(\frac{n_1 - r_1}{n_1}\right)\left[\left(\frac{n_2 - r_2}{n_2}\right)\right. \\ & D_2^+(r_1, n_1 + 1, r_2, n_2 + 1) \\ & \left. + \frac{r_2}{n_2} D_2^+(r_1, n_1 + 1, r_2 + 1, n_2 + 1)\right]. \end{aligned} \tag{4.1.6}$$

2. The value $n = \frac{N}{2}$ gives a special case of the optimal fixed sample size scheme; this will be discussed in section 4.2. The continuation risk at the origin in this case is given by

$$B_G^+(a_1, b_1, a_2, b_2) = n(C_1 + C_2) + \sum_{j=0}^{n} \sum_{k=0}^{n}$$

$$\frac{\binom{n}{j}\binom{n}{k}\beta(r_1, n_1 - r_1)\ \beta(r_2, n_2 - r_2)}{\beta(a_1, b_1 - a_1)\ \beta(a_2, b_2 - a_2)}$$

$$\times\ S(r_1, n_1, r_2, n_2). \qquad (4.1.7)$$

Note: for ease of notation, let $D_G^+(N)$ denotes the overall risk $D_G^+(0, 0, 0, 0)$ for OPT_2 and the special case where the group size is 2 will be denoted by $D_2^+(N)$.

4.1.2 Numerical results

In this section we present some numerical results to investigate the behaviour and the performance of the group sequential scheme. We have noticed that the changes in the loss constants K_1, K_2, sampling costs C_1, C_2 and the prior probabilities affect the optimal overall risk attainable using this scheme in a similar way to that in the fully sequential scheme under the various loss functions. Table (4.1) gives the results for $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$ with linear losses under two different priors and group size 2, the results are directly comparable with Table (3.7) in the last chapter. These loss constants, loss functions, costs of sampling and priors will be used throughout this chapter to illustrate the methods; the effect of departure from these conditions will be discussed. Again we notice the effect of change of priors on the overall risk $D_2^+(N)$ and the decreasing percentage reduction (DR%) in $D_2^+(N)$, gained by using the optimal maximum sample size, as N increases, suggesting truncation of the procedure could be useful. For example, a sample with N = 20 results in approximately 8.7% reduction in $D_2^+(N)$ under uniform priors and 7.5% under Beta priors (1, 3) v (1, 2). The optimal maximum sample size under this scheme is (82) using uniform priors (1, 2) v (1, 2) and (80) using Beta priors (1, 3) v (1, 2).

Table (4.1)

$D_2^+(N)$ as a function of N and (DR%) for OPT_2 with group size 2 where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear losses with two sets of priors.

	Be(1, 2) v Be(1, 2)			Be(1, 3) v Be(1, 2)	
N	$D_2^+(N)$	(DR%)	N	$D_2^+(N)$	(DR%)
10	32.0305	41.16	10	31.3524	31.03
20	26.9753	8.7	20	25.7253	7.5
30	25.3292	2.067	30	24.3937	1.95
40	24.9366	.485	40	24.0412	.47
50	24.8429	.107	50	23.9525	.10
60	24.8212	.0197	60	23.9324	.018
70	24.8170	.0028	70	23.9287	.0025
80	24.8164	.0004	80	23.9281	.0000
82	24.8163	.0000	82	23.9281	
84	24.8163				

Given the competing procedure OPT_1, we can study the efficiency of the group sequential scheme.

First, we test the robustness of the group sequential schemes with respect to group size. Considering the loss constants, loss function and priors as in Table (4.1), Table (4.2) gives the effect of group size on the relative efficiencies for a large sample size of N = 120.

Table (4.2)

Relative efficiencies (Eff%) of group sequential sampling for group sizes 2n, to purely sequential sampling OPT_1 for linear losses with $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, N = 120, with two sets of priors.

group size (2n)	priors	
	Be(1, 2) v Be(1, 2)	Be(1, 3) v Be(1, 2)
2	97.95	97.41
4	96.57	95.05
6	94.58	92.58
10	89.00	87.29
12	86.13	84.21
20	73.94	72.01
24	63.39	66.29
30	60.93	58.73
40	50.71	48.99
60	37.18	35.74
120	19.80	27.97

Note: For OPT_1, the overall risk D(120) = 24.3073 under uniform priors (1, 2) v (1, 2) and D(120) = 23.3073 under Beta priors (1, 3) v (1, 2).

The relative efficiency obviously decreases as the group size increases to, in the limit, the fixed sample size scheme, however, there is a small loss in efficiency for relatively large values of (2n), this would be tolerated since these schemes are easier to implement than the purely sequential scheme.

The effect of truncation on these group sequential schemes was investigated by considering the extreme case n = 1, where pair of observations are taken one at a time from both populations. The efficiencies relative to the corresponding truncated purely sequential design is given in Table (4.3).

As the value of N increases the relative efficiency of the group sequential design increases to a constant near N = 40 and reaches a quite respectable value for much smaller sample sizes, confirming that truncation is worth considering in this case, a limiting value of efficiency seems to exist.

Table (4.3)

Relative efficiencies of group sequential (n = 1) for linear losses with $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, varying N with two sets of priors.

N	Be(1, 2) v Be(1, 2)		Be(1, 3) v Be(1, 2)	
	overall risk of OPT_1	(Eff%) of OPT_2	overall risk of OPT_1	(Eff%) of OPT_2
10	32.9121	93.95	30.0599	95.88
20	26.1292	96.86	24.8118	96.45
30	24.7752	97.81	23.6925	97.12
40	24.4258	97.95	23.4026	97.34
50	24.3341	97.95	23.3283	97.39
60	24.3125	97.95	23.3112	97.39
70	24.3081	97.95	23.3079	97.40
80	24.3074	97.95	23.3073	97.40
90	24.3073	97.95	23.3073	
100	24.3073	97.95	23.3073	

4.2 Optimal fixed sample size scheme OFSS

In this section the optimal fixed sample size solution to the problem of selecting the better of two Binomial populations is developed. Fixed sample size means that exactly N observations are taken. However, N is partitioned into N_1 and N_2 (not necessarily equal), the number of observations taken from populations π_1 and π_2 respectively.

Based on N_1 and N_2 observations, to determine which decision is optimal at the point (r_1, n_1, r_2, n_2), where $r_i = a_i + c_i$, $n_i = b_i + N_i$, $0 \leq c_i \leq N_i$, $(i = 1, 2)$, we have to obtain the posterior expected losses of the decisions and the optimal decision is the decision associated with the smaller posterior expected decision loss (smaller risk).

4.2.1 Construction of the scheme OFSS

At the point (r_1, n_1, r_2, n_2), $n_1 + n_2 = b_1 + b_2 + N$, let S_1 and S_2 denote the stopping risks of taking decision D_1 and D_2 respectively. The termal decision rule for OFSS is as follows.

Take decision D_1 if $S_1 \leq S_2$.

Take decision D_2 if $S_1 > S_2$.

Where the stopping risk S_i $(i = 1, 2)$ for decision D_i is given by

$$S_i(r_1, n_1, r_2, n_2) = E\{L_i(D_i; p_1, p_2)\}$$

$$\pi(p_1, p_2 | r_1, n_1, r_2, n_2) \qquad (4.2.1)$$

where $L_i(D_i; p_1, p_2)$ is the decision loss for the decision D_i and $\pi(p_1, p_2 | r_1, n_1, r_2, n_2)$ is the posterior probability density of p_1 and p_2.

To compare this scheme with optimal sequential schemes we should find the prior risk of using it. At this stage we need to include the cost of sampling. Let $D_{FS}(N, N_1, N_2)$ be the prior risk of using OFSS (the average posterior expected loss), C_1, C_2 are the constant costs of sampling one observation from π_1, π_2 respectively, then at the point (r_1, n_1, r_2, n_2),

$$D_{FS}(N, N_1, N_2) = E\left[\min_{D_i}\left\{E(L_i(D_i; p_1, p_2) + N_1C_1 + N_2C_2)\right\}_{\pi(p_1, p_2 | r_1, n_1, r_2, n_2)}\right]$$

$$= \left[\sum_{c_1=0}^{N_1} \sum_{c_2=0}^{N_2} P(c_1, c_2) \min(S_1, S_2)\right] + N_1C_1 + N_2C_2 \tag{4.2.2}$$

where, $P(c_1, c_2)$, the predictive probability density function of c_1 and c_2, is given by

$$P(c_1, c_2) = \underset{\pi(p_1, p_2)}{E}\{f(c_1, c_2 | N_1, N_2, p_1, p_2)\}$$

$$= \frac{\binom{N_1}{c_1}\binom{N_2}{c_2}\beta(r_1, n_1 - r_1)\ \beta(r_2, n_2 - r_2)}{\beta(a_1, b_1 - a_1)\ \beta(a_2, b_2 - a_2)} \tag{4.2.3}$$

where $\pi(p_i)$, $f(c_i | N_i, p_i)$, $\beta(.,.)$ are defined as before.

If the prior probability distribution of p_1 and p_2 are uniform (that is Beta (1, 2) priors), then equation 4.2.3 becomes

$$P(c_1, c_2) = \begin{cases} \dfrac{1}{(N_1 + 1)(N_2 + 1)} & \text{if } N_1 \neq N_2 \\[2ex] \dfrac{1}{(n + 1)^2} & \text{if } N_1 = N_2 = n. \end{cases} \tag{4.2.4}$$

Using predictive probability (4.2.4) with a constant loss function, the formula (4.2.2) for $D_{FS}(N, N_1, N_2)$ will be reduced to the form

$$\frac{N_1!\,N_2!}{(N_1 + N_2 - 2)!} \sum_{c_1=0}^{N_1} \sum_{c_2=0}^{N_2} \min\left\{ \frac{K_1}{c_1!\,(N_1 - c_1)!} \sum_{j=c_2+1}^{N_2+1} \frac{(c_1 + j)!\,(N_1 + N_2 - c_1 - j + 1)!}{j!\,(N_2 + 1 - j)!}, \right.$$

$$\left. \frac{K_2}{c_2!\,(N_2 - c_2)!} \sum_{j=c_1+1}^{N_1+1} \frac{(c_2 + j)!\,(N_1 + N_2 - c_2 - j + 1)!}{j!\,(N_1 + 1 - j)!} \right\}$$

$$+ N_1C_1 + N_2C_2. \qquad (4.2.5)$$

As we mentioned in section (4.1), for $N_1 = N_2 = n = \frac{N}{2}$, OFSS is equivalent to the group sequential scheme with number of groups G = 1 since

$$D_{FS}(N, N_1, N_2) = n(C_1 + C_2) + \sum_{j=0}^{d_1} \sum_{k=0}^{d_2}$$

$$\frac{\binom{n}{j}\binom{n}{k} \beta(r_1, n_1 - r_1)\ \beta(r_2, n_2 - r_2)}{\beta(a_1, b_1 - a_1)\ \beta(a_2, b_2 - a_2)} \min(S_1, S_2)$$

$$\equiv B_G^+(a_1, b_1, a_2, b_2)$$

where $N_i = d_i = n$ $(i = 1, 2)$.

This optimal fixed sample size with equal allocation of observations will be denoted by OFSS(n) and its risk by $D_{FS}(N, n)$.

4.2.2 Numerical results

Some numerical work has been carried out to investigate the optimal fixed sample schemes OFSS and OFSS(n). We notice that loss constants, sampling costs and prior probabilities have the same influence on risks as seen in OPT_1 and OPT_2 schemes. Consider the same examples given in section (4.1.2) under uniform priors. Table (4.4) shows clearly that $D_{FS}(N, N_1, N_2)$ and $D_{FS}(N, n)$ decrease as N increases up to the optimal maximum sample size and then increase as N increases. This striking behaviour occurs only in these schemes since the rate of increase in sampling cost is larger than the rate of decrease in risk. The effect of sampling cost will be discussed in more details in section 4.3. In Table (4.4) the optimal sample size is 15 for OFSS and 16 for OFSS(n) under uniform priors.

Also, the same table displays the percentage reduction (DR%) relative to that for optimal sample size, which obviously decreases and then increases as N increases, for the above example with uniform priors, a sample with N = 10 results in approximately 7.97% increase in $D_{FS}(N, N_1, N_2)$.

Table (4.5) gives the best values of N_1 ($N_2 = N - N_1$) for various N under the same example above with two sets of priors. The best value of N_1 is defined as that value of N_1 which gives minimum overall risk for particular value of N. The program for finding the best values of N_1 and N_2 for linear losses, using formula (4.2.2) is given in appendix (4.1). With this program one can determine how many observations should be taken.

Table (4.4)

$D_{FS}(N, N_1, N_2)$ and $D_{FS}(N, n)$ as a function of N and the percentage reduction in risk (DR%) due to the optimal maximum sample size for $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear losses with uniform priors.

N	OFSS		N	OFSS(n)	
	$D_{FS}(N, N_1, N_2)$	(DR%)		$D_{FS}(N, n)$	(DR%)
10	36.1905	7.97	10	37.7778	9.44
15	33.5185	0.00	16	34.5185	0.00
20	34.5688	3.13	20	35.1515	1.83
30	40.1307	19.73	30	40.4167	17.09
40	47.7640	42.50	40	47.9365	38.87
50	56.2963	67.96	50	56.4103	63.42
60	65.2949	94.81	60	65.3763	89.39
70	74.5689	122.47	70	74.6296	116.20
80	84.0178	150.66	80	84.0650	143.54
90	93.5855	179.21	90	93.6232	171.23
100	103.237	208.00	100	103.268	199.17

Table (4.5)

The values of N_1 which give minimum $D_{FS}(N, N_1, N_2)$ in OFSS where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear loss function, for different values of N with two sets of priors.

N	priors	
	Be(1, 2) v Be(1, 2)	Be(1, 3) v Be(1, 2)
10	4	4,5
20	9,11	10
30	14,16	15
40	19,21	19
50	24,26	25
60	29,31	29
70	34,36	34
80	39,41	39,40
90	44,46	44,45
100	49,51	50

Regarding the same example discussed before under the two sets of priors Be(1, 2) v Be(1, 2) and Be(1, 3) v Be(1, 2), Table (4.6) shows that as the values of N increases the percent efficiency, (Eff%), of OFSS and OFSS(n) decreases. It seems from Table (4.6) that there is small differences in relative efficiencies for OFSS and OFSS(n) and these differences become smaller for large values of N. This suggests that it is reasonable to use the equal allocation scheme OFSS(n).

Table (4.6)

Relative efficiencies (Eff%) of optimal fixed sample size schemes OFSS, OFSS(n), to purely sequentially sampling OPT_1 for linear losses with $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$ with uniform priors and different values of N.

N	overall risk of OPT_1	(Eff%) of OFSS	(Eff%) of OFSS(n)
10	32.9121	90.94	87.12
20	26.1292	75.59	74.33
30	24.7752	61.74	61.30
40	24.4258	51.14	50.95
50	24.3341	43.23	43.14
60	24.3125	37.23	37.19
70	24.3081	32.60	32.57
80	24.3074	28.93	28.92
90	24.3073	25.97	25.96
100	24.3073	23.55	23.54

4.3 Discussion and conclusion

In this section we discuss the efficiencies of the group sequential and fixed sample size schemes relative to the fully sequential scheme. Table (4.7) seems to indicate that OPT_2 is more efficient than OFSS and OFSS(n). It is clear from the table that as N increases the relative efficiencies of the group sequential design increases until reaches a maximum of efficiency near N = 40. Note that the value of $D_2^+(20)$ is 108.18% of the value of $D_2^+(40)$. This behaviour confirms that truncation is worth considering in this case. On the other hand the relative efficiency of fixed sample size sampling decreases; however, it is worth considering if there is a high degree of truncation. The insignificant differences in the relative efficiencies of OFSS and OFSS(n) suggests the use of equal allocation scheme OFSS(n) in fixed sample size procedures.

The comparison above assumes that observation cost in all cases remain the same whatever sampling is being used, in practice, an adjustment in cost may be necessary to reflect the ease of use of some of the sampling methods. In practice the factors such as the cost of sampling, ethical considerations, delayed and instantaneous responses etc. may play important roles in choosing the sampling method.

Under the assumption of equal sampling cost and other related factors the fully sequential scheme with a maximum total sample size of N will have a smaller D(N) than $D_2^+(N)$ for group sequential scheme and $D_{FS}(N, N_1, N_2)$ and $D_{FS}(N, n)$ for fixed sample size schemes with a fixed sample size N. Suppose that there is a fixed cost associated with each sample, in addition to a cost per unit sampled, where the fixed cost is the same irrespective of the sample size. In the case of fixed sample schemes of N observations, the fixed cost is incurred once. For OPT_1 each observation is considered as a separate sample, the fixed cost may be incurred up to N times, whilst in OPT_2 is incurred G times (G is the number of groups in OPT_2 such that N = 2nG). In sampling from a production process for example, it may be necessary to stop the process

Table (4.7)

Relative efficiencies of group sequential scheme (n = 1) and fixed sample size schemes OFSS, OFSS(n) for linear losses with $K_1 = K_2 = 1000$, $C_1 = C_1 = 1$, uniform priors with various N.

N	Risk for OPT_1	Eff(%) of OPT_2	Eff(%) of OFSS	Eff(%) of OFSS(n)
10	32.9121	93.95	90.94	87.12
20	26.1292	96.86	75.59	74.33
30	24.7752	97.81	61.74	61.30
40	24.4258	97.95	51.14	50.95
50	24.3341	97.95	43.23	43.14
60	24.3125	97.95	37.23	37.19
70	24.3081	97.95	32.60	32.57
80	24.3074	97.95	28.93	28.92
90	24.3073	97.95	25.97	25.96
100	24.3073	97.95	23.55	23.54

in order to take a sample at a particular point in the process and a fixed cost of stopping the process and starting it again may be incurred. If the situation requires stopping the process after each item produced as in OPT_1 (or after a group of items) as in OPT_2 then the total cost of sampling will be greatest with OPT_1 then OPT_2 than with OFSS and OFSS(n).

However, the benefits gained from using OPT_1 and OPT_2 as measured by the overall risk of sampling may be considerably greater than the benefits of the OFSS and OFSS(n), in this context OPT_1 will be superior to both OPT_2 and fixed sample size schemes. Generally speaking, if the observations are very costly and no fixed cost associated with the sampling so that the cost of sampling is a function of the observations only, then OPT_1 and OPT_2 are preferable. On the other hand, if the fixed cost associated with sampling stage is the most important and not the cost of observations then the optimal fixed sample schemes may be preferred. The very slight loss of efficiency of OPT_2 will usually be more compensated for by its greater simplicity of use comparable with OPT_1 in terms of time and computer storage required to output sampling scheme.

CHAPTER 5

BAYESIAN SEQUENTIAL SUBOPTIMAL SCHEMES FOR SELECTING THE BETTER OF TWO BINOMIAL POPULATIONS

5.0 Introduction and summary

The classical approach does not take into account any information that the experimenter may have about the quality of a product or drug before procedures start. Moreover, there are some analytical and computational difficulties associated with the use of the optimal schemes. These motivations are behind the attempt to find and investigate some procedures which are simple to operate and which include the use of prior information. In this chapter we propose and study some suboptimal schemes for selecting the better of two Binomial populations.

The construction of the chapter is as follows.

In section 5.1 we study and investigate the suboptimal schemes LA.

In section 5.2 we propose and investigate some suboptimal schemes depending on the posterior estimates of p_1 and p_2.

Section 5.3 contains the risk performance of the suboptimal schemes given in section 5.2.

A brief discussion about these schemes is given in section 5.4.

5.1 Description of Look Ahead (LA) schemes

These schemes consider decisions in a forward sense rather than working backwards using dynamic programming. The 1-step look aheaded (OLA) scheme proceeds as if sampling stops after the next trial. The 2n-step look aheaded (GLA) proceeds as if sampling stops after the next stage where block of 2n, n on each population, observations are taken at a time. Let N be the maximum number of observations such that $G = \frac{N}{2n}$, where G is the number of groups. Again, consider the decisions

$$D_1: p_1 \leq p_2$$

and

$$D_2: p_1 > p_2.$$

In the following we describe the schemes OLA and GLA.

5.1.1 The scheme OLA

At the point (r_1, n_1, r_2, n_2), let

S_1 and S_2 denote the stopping risks of taking decision D_1 and D_2 respectively, calculated as before,

B_1^* denotes the continuation risk for π_1

B_2^* denotes the continuation risk for π_2

D^* denotes the minimum risk attainable using OLA.

Then,

$$B_1^* = C_1 + \frac{r_1}{n_1} S(r_1 + 1, n_1 + 1, r_2, n_2) + \frac{n_1 - r_1}{n_1}$$

$$S(r_1, n_1 + 1, r_2, n_2)$$

and

$$B_2^* = C_2 + \frac{r_2}{n_2} S(r_1, n_1, r_2 + 1, n_2 + 1) + \frac{n_2 - r_2}{n_2}$$

$$S(r_1, n_1, r_2, n_2 + 1)$$

where

$$S = \min(S_1, S_2).$$

The stopping rule is as follows:

Stop sampling if $\min(B_1^*, B_2^*) \geqslant S$

Continue sampling if $\min(B_1^*, B_2^*) < S$.

(We dropped the argument for ease of notation.)

The terminal decision rule is as follows:

Take D_1: $p_1 \leqslant p_2$ if $D^* = S_1$

Take D_2: $p_1 > p_2$ if $D^* = S_2$.

It is likely that $\min(B_1^*, B_2^*) > S$ at the origin since the results in section 3.5 give

$B_1^* = C_1 + S_1$ if both transition points are S_1 points

and

$B_2^* = C_2 + S_2$ if both transition points are S_2 points.

It is not known what happens when the transition points are S_1 and S_2 but some numerical work on this scheme produced sampling rules with no observations; hence these schemes are not considered further for single observations.

5.1.2 The scheme GLA

At the point (r_1, n_1, r_2, n_2), let

S_1, S_2 defined as before,

B^* denotes the continuation risk with both populations,

D_G^* denotes the minimum risk attainable using GLA.

Then,

$$B^* = n(C_1 + C_2) + \sum_{j=0}^{n} \sum_{k=0}^{n} E\left\{\binom{n}{j}\binom{n}{k} p_1{}^j q_1{}^{n-j} p_2{}^k q_2{}^{n-k}\right\}$$

$$S(r_1 + j, n_1 + n, r_2 + k, n_2 + n)$$

where

$$S = \min(S_1, S_2).$$

The stopping rule is:

Stop sampling if $B^* \geq S$

Continue sampling if $B^* < S$.

Now, if we wish to compare the risk performance of this scheme with respect to the optimal scheme OPT_2 then the required minimum risk D_G^* at the point (r_1, n_1, r_2, n_2) is calculated as follows.

$$D_G^* = \begin{cases} S_1 & \text{if } S_1 = \min(S_1, S_2, B^*) \\ S_2 & \text{if } S_2 = \min(S_1, S_2\ B^*) \\ n(C_1 + C_2) + \sum_{j=0}^{n} \sum_{k=0}^{n} E\left\{ \binom{n}{j} \binom{n}{k} p_1^{\,j} q_1^{\,n-j} p_2^{\,k} q_2^{\,n-k} \right\} & \\ D_G^*(r_1 + j, n_1 + n, r_2 + k, n_2 + n) & \\ & \text{if } B^* = \min(S_1, S_2, B^*). \end{cases}$$

The terminal decision rule is:

Take decision D_1 if $D_G^* = S_1$

Take decision D_2 if $D_G^* = S_2$.

Note for $n = 1$, the scheme GLA is reduced to the special case, denoted by TLA, where 2 observations are taken at a time, one from each population. The minimum risk for TLA is denoted by D_2^*.

Some numerical results concerning the efficiency of GLA relative to OPT_2 are given in Tables (5.1 - 5.2). Let the set of values $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$ with uniform priors is denoted by SET_1.

Table (5.1) shows the relative efficiencies of GLA under various group sizes 2n, where $N = 120$ and the set of values SET_1. The scheme is nearly fully efficient for relatively small group size; as expected the efficiency reaches 100% for large group size since this would imply an overall large number of observations, probably larger than the optimal maximum number of observations.

Table (5.1)

Relative efficiencies (Eff%) of GLA for group size 2n, to OPT_2 for linear losses with $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$ with uniform priors and N = 120.

2n =	2	4	6	10	12	20	24
(Eff%) =	64.76	91.49	97.94	98.95	99.85	99.90	100.00

2n =	30	40	60	120
(Eff%) =	100.00	100.00	100.00	100.00

Table (5.2) demonstrates the relative efficiencies of TLA under various values of N using SET_1 values. It is clear from this table that as N increases the relative efficiency decreases up to a constant after which there is no change in the efficiency of the scheme TLA.

Table (5.2)

Relative efficiencies of TLA where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, and uniform priors with N = 10(10)100 under linear losses.

N =	10	20	30	40	50	60
(Eff%) =	85.26	69.93	66.05	65.07	64.83	64.77

N =	70	80	90	100
(Eff%) =	64.76	64.76	64.76	64.76

5.2 Description of the schemes δ_1 and δ_G

In this section some Bayesian suboptimal schemes are proposed with decision criteria based on the posterior probabilities of p_1 and p_2. These are prompted by the need for a quick, easy schemes, to select the better of two

Binomial populations, which allow for the incorporation of prior information about the populations with sampling information, ignoring the decision-theoretic structure and the indifference-zone formulation.

Suppose that observations are taken sequentially one at a time or group at a time, with a maximum sample size of N, and are assumed to be independent with probability

$$f(c_i) = \binom{d_i}{c_i} p_i^{c_i} (1 - p_i)^{d_i - c_i}, \qquad i = 1, 2,$$

where p_i is the probability of success for π_i. Further, assume that p_i is assigned a Beta prior distribution with integer parameters a_i, b_i (or $Be(a_i, b_i)$) with density function proportional to

$$p_i^{a_i - 1} q_i^{b_i - a_i - 1}, \qquad 1 \leq a_i \leq b_i - 1, \qquad q_i = 1 - p_i$$

then after c_i successes in d_i trials on π_i ($i = 1, 2$), the posterior distribution of p_i is Beta with integer parameters r_i, n_i where $r_i = a_i + c_i$ and $n_i = b_i + d_i$ and the posterior expectation of p_i or the predictive probability that the next trial results in a success is $\frac{r_i}{n_i}$, $i = 1, 2$. These Bayesian sequential suboptimal schemes, termed δ-schemes with δ_1 refers to the fully sequential scheme and δ_G to the group sequential scheme, are based merely on

$$\delta = \frac{r_1}{n_1} - \frac{r_2}{n_2} \qquad (5.2.1)$$

and a preassigned constant δ_0 ($0 \leq \delta_0 \leq 1$) serving as an appropriate distance measure of the differences between the populations.

5.2.1 Formulation of δ_1

As we mentioned, in these schemes we sample observations sequentially one at a time ignoring any loss structure or costs of conducting the schemes. The difference δ is used both to determine the type of observation and the stopping rules and the terminal decision rules. For all schemes D_1 is made at termination, including N, if $\delta \leq 0$ and D_2 if $\delta > 0$ and for all sequential schemes sampling is stopped once $|\delta| > \delta_0$.

Let the stopping and terminal decision rules for these schemes are denoted by DS and DT respectively.

The fully sequential scheme is considered under the following six sampling rules A_1 - A_6, where $|\delta| \leq \delta_0$.

A_1: sample from population 1 (π_1) at next trial if $\delta < 0$.
sample from population 2 (π_2) at next trial if $\delta > 0$.
sample at random from (π_1, π_2), if $\delta = 0$.

A_2: A modification of A_1 to break ties, there are unlikely to be ties for non-integer prior parameters.
π_1 if $\delta < 0$ or $\delta = 0$ and $\lambda = n_1 - n_2 < 0$.
π_2 if $\delta > 0$ or $\delta = 0$ and $\lambda > 0$.
(π_1, π_2) if $\delta = 0$ and $\lambda = 0$.

A_3: π_1 if $\delta < 0$ or $\delta = 0$ and $\lambda > 0$.
π_2 if $\delta > 0$ or $\delta = 0$ and $\lambda < 0$.
(π_1, π_2) if $\delta = 0$ and $\lambda = 0$.

A_4: π_1 if $\delta > 0$.

π_2 if $\delta < 0$.

(π_1, π_2) if $\delta = 0$.

A_5: π_1 if $\delta > 0$ or $\delta = 0$ and $\lambda < 0$.

π_2 if $\delta < 0$ or $\delta = 0$ and $\lambda > 0$.

(π_1, π_2) if $\delta = 0$ and $\lambda = 0$.

A_6: π_1 if $\delta > 0$ or $\delta = 0$ and $\lambda > 0$.

π_2 if $\delta < 0$ or $\delta = 0$ and $\lambda < 0$.

(π_1, π_2) if $\delta = 0$ and $\lambda = 0$.

Rules (A_1 - A_3) are mainly concerned with the continuation of sampling with the population which has the inferior probability of success. Rules (A_4 - A_6) are mainly concerned with the continuation of sampling with the population which has higher probability of success.

It may seem illogical to choose the population with the smaller posterior expectation (A_1 - A_3) for the next observation but it should be noted that although this is likely to produce a smaller number of successes than A_4 - A_6 it may give a better discrimination between the two populations. The use of n_i to break ties has been used to give improved results for the suboptimal schemes in the two armed bandit problem in Jones and Kandeel (1984).

A comparison with OPT_1 may be carried out but it should be noted that the proposed scheme does not depend on loss constants and costs and where they are very different for the two decisions and for sampling for the two populations the comparison is not sensible. Hence in the numerical comparisons given later equal loss constants and costs are assumed. If the risk of using this scheme is near to that of the optimal then this suggests a way of displaying the optimal decisions as a series of points where the suboptimal decision differs from the optimal decision.

For the purpose of risk comparison with OPT_1, we calculate the stopping and continuation risks for these schemes as follows.

At the point (r_1, n_1, r_2, n_2), let

S_1 and S_2 denote the stopping risks of taking decisions D_1 and D_2 respectively and calculated as before

B_1' denotes the continuation risk with π_1

B_2' denotes the continuation risk with π_2

D' denotes the minimum risk attainable using this scheme.

Then,

$$D' = \begin{cases} S_1 & \text{if } |\delta| > \delta_0 \text{ and } \delta \leq 0 \\ S_2 & \text{if } |\delta| > \delta_0 \text{ and } \delta > 0 \\ B_1' = C_1 + \dfrac{r_1}{n_1} D'(r_1 + 1, n_1 + 1, r_2, n_2) + \dfrac{n_1 - r_1}{n_1} & \\ \qquad D'(r_1, n_1 + 1, r_2, n_2) & \text{if } |\delta| \leq \delta_0 \text{ and } \delta \leq 0 \\ B_2' = C_2 + \dfrac{r_2}{n_2} D'(r_1, n_1, r_2 + 1, n_2 + 1) + \dfrac{n_2 - r_2}{n_2} & \\ \qquad D'(r_1, n_1, r_2, n_2 + 1) & \text{if } |\delta| \leq \delta_0 \text{ and } \delta > 0. \end{cases}$$

5.2.2 Formulation of δ_G

This scheme is similar to δ_1 as it depends on the posterior estimates of the parameters p_1 and p_2. The observations here are taken sequentially in groups of 2n observations at each stage, n on each population until a decision is reached or N is reached.

The stopping and terminal decision rules used are DS and DT described in subsection 5.2.1. Here there is no choice of sampling between the populations as we continue sampling with both populations if $|\delta| \leqslant \delta_0$ is satisfied, otherwise we stop sampling and proceeds to the terminal decision rule DT. If the sampling has not stopped before N according to this stopping rule then it should be terminated at N.

Using the same justification as that given for δ_1 for comparing it with OPT_1, a risk comparison of δ_G with OPT_2 may be carried out and the various risks are calculated as follows.

At the point (r_1, n_1, r_2, n_2), let
S_1 and S_2 are defined and calculated as before,
B' be the continuation risk with both populations,
D_G' be the optimal risk using this scheme.
Then,

$$D_G' = \begin{cases} S_1 & \text{if } |\delta| > \delta_0 \text{ and } \delta \leqslant 0 \\ S_2 & \text{if } |\delta| > \delta_0 \text{ and } \delta > 0 \\ B' = n(C_1 + C_2) + \sum_{j=0}^{n} \sum_{k=0}^{n} E\left\{\binom{n}{j}\binom{n}{k} p_1^{\,j} q_1^{\,n-j} p_2^{\,k} q_2^{\,n-k}\right\} & \\ \qquad D_G'(r_1 + j, n_1 + n, r_2 + k, n_2 + n) & \text{if } |\delta| \leqslant \delta_0. \end{cases}$$

5.3 Risk performance of δ-schemes

5.3.1 Numerical results for δ_1

Tables (5.3 - 5.5) contain some numerical results about the schemes $\delta_1(A_1)$, ..., $\delta_1(A_6)$ using the set of values SET_1. It is clear from Table (5.3) that the optimal overall risk using δ_1, denoted by $D'(N)$, decreases then increases as δ_0 increases with the rate of decrease greater than the rate of increase and $D'(N)$ may take constant value for very large values of δ_0 (say, $\delta_0 > .7$). The same table indicates that the scheme $\delta_1(A_5)$ is the best among δ_1-schemes, then $\delta_1(A_2)$ comes next. It has been found that there exists an optimal value for δ_0 for each scheme which gives minimum overall risk for that scheme. For the above example all schemes have minimum overall risks around $\delta_0 = .3$. Table (5.4) demonstrates the same behaviour for various values of N using $\delta_1(A_5)$ scheme. Table (5.5), where linear losses, SET_1, different schemes, $\delta_0 = .3$ with different values of N are used, shows that there exists an optimal maximum sample size for each scheme and may vary as δ_0 varies. Table (5.6) presents the efficiency of the scheme $\delta_1(A_5)$ relative to OPT_1 using the set of values SET_1 and $\delta_0 = .3$; it shows clearly that the relative efficiency decreases as N increases.

Table (5.3)

D'(N) as a function of δ_0 for the suboptimal schemes $\delta_1(A_1)$, ..., $\delta_1(A_6)$ when $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, $N = 50$, linear losses with uniform priors (Be(1, 2) v Be(1, 2).

	The suboptimal schemes					
δ_0	$\delta_1(A_1)$	$\delta_1(A_2)$	$\delta_1(A_3)$	$\delta_1(A_4)$	$\delta_1(A_5)$	$\delta_1(A_6)$
0.0	84.3333	84.3333	84.3333	84.3333	84.3333	84.3333
0.1	84.3333	84.3333	84.3333	84.3333	84.3333	84.3333
0.2	47.0947	44.9263	47.1059	47.2358	44.9850	47.2455
0.3	46.2999	43.3726	46.3770	46.7034	43.2366	46.7033
0.4	54.4266	51.6936	54.6630	55.0861	52.5115	55.3294
0.5	66.3443	64.3450	66.5951	65.8061	63.6974	66.0809
0.6	71.0941	68.8301	71.3478	71.3249	69.0031	71.6039
0.7	72.6156	70.3550	72.8743	72.8433	70.5269	73.1306
0.8	72.6181	70.3575	72.8768	72.8458	70.5294	73.1331
0.9	72.6181	70.3575	72.8768	72.8458	70.5294	73.1331
1.0	72.6181	70.3575	72.8768	72.8458	70.5294	73.1331

Table (5.4)

D'(N) as a function of δ_0 for the scheme $\delta_1(A_5)$, when $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear losses with uniform priors and various values of N.

δ_0	N				
	10	20	30	40	50
0.0	84.3333	84.3333	84.3333	84.3333	84.3333
0.1	84.3333	84.3333	84.3333	84.3333	84.3333
0.2	46.4716	44.4298	44.2351	44.5503	44.9850
0.3	42.5384	39.3985	39.8048	41.3839	43.2366
0.4	42.7972	41.1905	43.7961	47.9748	52.5115
0.5	43.5979	44.0982	40.2719	56.2638	63.6974
0.6	44.1138	45.7989	52.1678	60.3530	69.0031
0.7	44.1138	45.9794	52.7200	61.3777	70.5269
0.8	44.1138	45.9794	52.7200	61.3777	70.5294
0.9	44.1138	45.9794	52.7200	61.3777	70.5294

Table (5.5)

D'(N) as a function of N for suboptimal schemes $\delta_1(A_1)$, ..., $\delta_1(A_6)$, when $\delta_0 = 0.3$, $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear losses with uniform priors.

N	suboptimal schemes					
	$\delta_1(A_1)$	$\delta_1(A_2)$	$\delta_1(A_3)$	$\delta_1(A_4)$	$\delta_1(A_5)$	$\delta_1(A_6)$
10	45.7039	42.7552	45.7370	45.8420	42.5384	45.8584
20	42.7307	39.9549	42.8050	42.6165	39.3985	42.6536
30	43.1141	40.2993	43.2030	43.0904	39.8048	43.1291
40	44.5216	41.7041	44.6000	44.6987	41.3839	44.7155
50	46.2999	43.3726	46.3770	46.7034	43.2366	46.7033
60	48.2236	45.1415	48.3070	48.8614	45.1866	48.8544
70	50.1914	46.9930	50.2668	51.0800	47.2405	51.0505
80	52.2473	48.9027	52.3222	53.3949	49.3562	53.3498
90	54.3014	50.8065	54.3766	55.7130	51.4694	55.6528
100	56.3937	52.7415	56.4688	58.0571	53.6026	57.9817

Table (5.6)

Efficiencies of $\delta_1(A_5)$, relative to OPT_1, where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear losses, $\delta_0 = .3$, uniform priors with different values of N.

N =	10	20	30	40	50	60
(Eff%) =	77.37	66.32	62.24	59.02	56.28	53.80

N =	70	80	90	100
(Eff%) =	51.46	49.25	47.23	45.35

5.3.2 Numerical results for δ_G

This subsection contains some numerical results about the efficiency of δ_G relative to OPT_2 using the set of values SET_1. Table (5.7) shows, for $\delta_0 = .4$, $N = 120$ with different group sizes 2n and the set of values SET_1, the effect of grouping on the efficiency of δ_G scheme.

Table (5.7)

Efficiencies of δ_G relative to OPT_2, where $\delta_0 = .4$, $N = 120$, $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear losses and uniform priors with different group sizes 2n.

2n =	2	4	6	10	12	20
(Eff%) =	37.52	36.45	34.19	38.09	35.58	40.58

2n =	24	30	40	60	120
(Eff%) =	41.47	45.62	51.73	64.49	100.00

We notice from Table (5.7) that as the group size 2n increases the efficiency decreases very slowly then increases rapidly. This suggests that there are some values of 2n where the efficiencies of δ_G attain minimum values.

Table (5.8) shows the efficiencies of δ_2, where 2 observations are taken at a time, $\delta_0 = .4$, the set of values SET_1 with various values of N, relative to OPT_2. We note that the relative efficiency decreases as N increases.

Table (5.8)

Efficiencies of δ_2 relative to OPT_2, where $\delta_0 = .4$, $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, $n = 1$, linear losses with uniform priors and different values of N.

N =	10	20	30	40	50	60	70
(Eff%) =	97.35	91.10	83.55	75.97	68.94	62.72	57.31

N =	80	90	100
(Eff%) =	52.66	48.64	45.15

5.4 Discussion

From Table (5.2) and Table (5.8) the relative efficiencies of TLA and δ_2 decreases as N increases. However using TLA gives a rate of decrease less than δ_2. We notice from Table (5.1) that GLA has remarkable increase in efficiency relative to OPT_2 as n increases. The same behaviour is noted in δ_2 but it is less efficient than GLA (see Table (5.7)). The schemes GLA and δ_G with group size $2n = N$ have exactly the same performance as OPT_2 since all N observations are taken. GLA and δ_G are simple to use and have a good performance hence it suggests that they should be used in practice particularly for moderate group sizes.

CHAPTER 6

MONTE CARLO STUDIES OF BINOMIAL OPTIMAL SCHEMES OPT_1, OPT_2 AND OFSS

6.0 Introduction and summary

For some applications it is of interest to compare the schemes mentioned in the previous chapters under criteria other than overall risk. If the schemes are to be used in clinical trials, measures such as the probability of correctly selecting the better treatment at termination or the expected number of trials on the poorer treatment are of more use. To calculate these and other measures, some Monte Carlo (MC) simulations were carried out to assess the effectiveness of our procedures. For the purpose of comparing the procedures under these criteria, we have categorized them into two groups. The first group, where K_1, K_2, C_1, C_2, loss function and prior information are involved, includes the optimal schemes OPT_1, OPT_2 and OFSS. The second group includes all suboptimal schemes, where K_1, K_2, C_1, C_2 and loss function are ignored, together with some other procedures given in the literature.

Chapter 6 is devoted to investigating the performance characteristics of the first group, while chapter 7 is devoted to studying the performance characteristics of the second group.

The contents of chapter 6 can be summarized as follows.

The Monte Carlo study is described in section 6.1.

Section 6.2 explains how to implement the optimal schemes and describes the stages by which the calculations of the performance characteristics of these schemes may be carried out.

The results of MC estimates and discussion are given in section 6.3.

We conclude this chapter by some remarks given in section 6.4.

6.1 Description of the MC studies

In this section we briefly describe the method of MC simulation as it is applied to our procedures. Monte Carlo studies have been carried out to investigate some of the performance characteristics of the proposed procedures. Computer programs, which simulate the operations of these procedures, were written in FORTRAN and run at University of Manchester Regional Computer Centre (UMRCC) on CYBER 205.

The simulation programs perform a large number of runs t (t = 10,000), which are assumed to be independent, in order to obtain MC estimates with high precision. At each run mutually independent Bernoulli observations are generated by using the assumed probability model with p_1 and p_2 specified in advance and then the selection procedure is applied. The observed values of several performance measures are accumulated. At the end of all runs, these accumulated values are divided by t to obtain the MC estimates of the performance characteristics of interest.

The subroutine Gϕ5CAF of NAG (Numerical Algorithm Group)

Library (NAGFLIB 1977), available at UMRCC, is used to generate a uniform variate x ($0 \leq x < 1$). Population π_i, with probability of success p_i, scored success if corresponding random number $x < p_i$ and failure if $x \geq p_i$ ($i = 1, 2$). Formally,

If X is uniformally distributed random variable over (0, 1), that is $X \sim U(0, 1)$, then

$$P(X < p_i) = \int_0^{p_i} dx = p_i = P(Y = 1) \qquad (6.1.1)$$

and

$$P(X \geq p_i) = 1 - p_i = P(Y = 0), \qquad 0 \leq p_i \leq 1$$

where Y is a random variable with value 0 or 1.

A Binomial B(d, p) random variable C can be written as $C = \sum_{j=1}^{d} Y_j$, where Y_j are independent Bernoulli random variables, each taking the values $Y_j = 1$ with probability p or $Y_j = 0$ with probability $(1 - p)$. Thus, to simulate such a C, we need just simulate d independent U(0, 1) random variables, $U_1, U_2, \ldots, U_d$ and set $Y_j = 1$ if $U_j < p$ and $Y_j = 0$ if $U_j \geq p$.

The values of p_1 and p_2 can be specified as follows:

I - Fixed p_1, p_2, where for each run of 10000 trials the same p_1 and p_2 are used.

II - Generated p_1, p_2, where the pair of values changes for each trial using one of the following methods:

(1) They are generated from uniform distribution using GΦ5CAF random number generator if they have

independent uniform priors over (0, 1).

(2) They are generated from Beta distribution using Gϕ5DLF random number generator if they have independent Beta priors.

With the observed value p_i using either step (1) or (2), the values of Y can be considered as the observed values of a random variable, possessing the Bernoulli distribution that should be simulated. According to the sampling schemes, the following quantities are required for input.

(i) For the optimal schemes OPT_1, OPT_2, OFSS:
K_1, K_2, C_1, C_2, priors, loss function, N.

(ii) For the suboptimal schemes:
N, priors, δ_0.

(iii) For fixed sample size FSS:
N, priors.

As measures of performance of the proposed procedures we shall use the following quantities.

(a) P(CS): Probability of correct selection.

In a MC experimentation the population that has the greatest probability of success is known to us, so we can check if the procedure gives a correct selection. After t repetitions we estimate the probability of correct selection by the fraction of correct selections in the t repetitions. It can be computed as follows.

$P(D_i/D_i)$ = The proportion of number of times when the procedure stops and takes decision D_i given decision D_i is true in t repetitions,

$$P(CS) = \sum_{i=1}^{2} P(D_i/D_i), \text{ where } D_1\text{:}\ \ p_1 \leqslant p_2,$$
$$D_2\text{:}\ \ p_1 > p_2. \qquad (6.1.2)$$

(b) $E(M)$: Expected sample size, where M denotes the actual number of observations taken from the given population.

An estimate of $E(M)$ is given by

$$E(M) = \sum_{j=1}^{t} M_j/t, \qquad (6.1.3)$$

where M_j denotes the number of observations taken from both populations in the j^{th} run. $E(M)$ is not given for fixed sample size schemes since $E(M) = N$.

(c) $E(N_{(1)})$: Expected number of observations on the inferior population.

An estimate of $E(N_{(1)})$ is given by

$$E(N_{(1)}) = \sum_{j=1}^{t} N_{(1)j}/t, \qquad (6.1.4)$$

where $N_{(1)j}$ is the number of observations assigned to the inferior population in the j^{th} run. $E(N_{(1)})$ is not given for the group sequential schemes since $E(N_{(1)}) = E(M)/2$ and for the fixed sample size schemes since $E(N_{(1)}) = N/2$.

(d) $E(N^*_{(1)})$: Expected number of observations on the inferior population if sampling continues with chosen population for the remaining $(N - M)$ observations.

A reasonable approximation, which always overestimates the true values, can be obtained by using

$$E(N_{(1)}) + [(N - E(M))(1 - P(CS))]. \qquad (6.1.5)$$

(e) $E(R)$: Expected number of successes.

An estimate of $E(R)$ is given by

$$E(R) = \sum_{j=1}^{t} (R_1 + R_2)_j/t, \qquad (6.1.6)$$

where $(R_1 + R_2)_j$ is the number of successes gained in the j^{th} run with R_i successes from population π_i $(i = 1, 2)$. For fixed p_1 and p_2 we have

$$E(R) = p_{[2]}\, E(M) + (p_{[1]} - p_{[2]})\, E(N_{(1)}), \qquad (6.1.7)$$

where $p_{[1]} = \min[p_1, p_2]$ and $p_{[2]} = \max[p_1, p_2]$, under sequential schemes and $E(R) = N/2$ under fixed sample size schemes.

(f) $E(R^*)$: Expected number of successes if sampling continues with the chosen population for the remaining $(N - M)$ observations.

For fixed p_1, p_2, this measure can be calculated by

$$E(R^*) = E(R) + [N - E(M))]p_{[2]} + [E(N^*_{(1)}) - E(N_{(1)})](p_{[1]} - p_{[2]}). \qquad (6.1.8)$$

6.2 Description of the computation technique

This section describes the stages of computation used to carry out the MC studies of OPT_1, OPT_2 and OFSS and produce the results given in the next section. The listings of the computer programs are made available in appendices (6.1 - 6.3) for further analysis of the procedures.

We can not actually use a direct method (that is, start from the origin and follow the reachable points) to perform the simulation for the optimal schemes as we face the problem of storing the whole set of optimal points. Therefore, we propose an alternative method to overcome this problem. This method involves the use of the suboptimal schemes δ_1 (δ_G) in such a way that we deal only with the exceptional points, i.e. where OPT_1 (OPT_2) differs from δ_1 (δ_G). It consists of the following stages:

Stage 1:

In this stage we compare OPT_1 (OPT_2) with δ_1 (δ_G) and display the points where OPT_1 (OPT_2) gives a different decision from δ_1 (δ_G) for particular value of δ_0. That is, we cut off the points where OPT_1 (OPT_2) and δ_1 (δ_G) coincide and output the exceptional points. We choose the value of δ_0 which gives minimum number of exceptional points. This is one way of investigating the effect of δ_0 in the suboptimal schemes δ_1 and δ_G, where the best value of δ_0 is that which gives mimimum number of exceptional points and makes the proposed suboptimal schemes nearest to the optimal scheme. These exceptional points with their types S_1, S_2, B_1, or B_2 in the case of OPT_1 (S_1, S_2 or B in the case of OPT_2) and the number of them will be saved in a file to be used in the next stage.

It should be pointed out that the number of exceptional points grows rapidly with N. Therefore, a considerable storage capacity is required, particularly for large N.

Some numerical examples are presented Tables (6.A_1, 6.A_2, 6.B_1 and 6.B_2) to display the effect of δ_0 on the number of exceptional points and the best values of δ_0 for OPT_1 and OPT_2 (n = 1), where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear loss function with uniform priors.

It is worth mentioning that the risk performance of δ-schemes given in Chapter 5, using the above example, suggests that the best value of δ_0 which gives minimum risk is around 0.3, whereas the values giving the minimum number of exceptional points is around 0.1. This may be due to the sampling procedure as there are many points which are not reached during the sampling.

Table (6.A_1) shows that N = 50 generates 316251 points in four dimensional integer space for OPT_1, over 82% of which give the same decision as δ_1 with $\delta_0 = 0.1$; in addition, Table (6.B_1) shows that N = 50 generates 6201 points for OPT_2 (n = 1), over 90% of which give the same decision as δ_2. Symmetry could be used to make a further reduction of these exceptional points when the loss and cost constants are the same for each decision.

Table (6.A_1)

The number of exceptional points obtained as a results of comparison between OPT_1 and δ_1 (A_5) schemes, where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, $N = 50$, uniform priors under linear loss function.

δ_0	No. of exceptional pts.	No. of coincident pts.
0.0	59927	256324
0.1	54762	261489
0.2	88493	227758
0.3	128922	187329
0.4	166491	149760
0.5	199286	116965
0.6	224580	91671
0.7	243468	72783
0.8	255577	60674
0.9	261058	55193
1.0	261510	54741

Table (6.A_2)

The best values of δ_0, giving minimum number of exceptional points, that used in comparison between OPT_1 and δ_1 (A_5) schemes for different values of N where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, uniform priors under linear loss function.

N	=	10	20	30	40	50
δ_0	=	0.2	0.1	0.1	0.1	0.1

Table (6.B_1)

The number of exceptional points obtained as a results of comparison between OPT_2 (n = 1) and δ_2 schemes, where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, N = 50, uniform priors under linear loss function.

δ_0	No. of exceptional pts.	No. of coincident pts.
0.0	368	5833
0.1	580	5621
0.2	1460	4741
0.3	2291	3910
0.4	3057	3144
0.5	3708	2493
0.6	4142	2059
0.7	4494	1707
0.8	4732	1469
0.9	4840	1361
1.0	4852	1349

Table (6.B_2)

The best values of δ_0, giving minimum number of exceptional points, that used in comparison between OPT_2 (n = 1) and δ_2 schemes for different values of N, where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, uniform priors under linear loss function.

N	=	10	20	30	40	50
δ_0	=	0.3	0.2	0.1	0.0	0.0

Stage 2:

The data (the exceptional set of points) produced in Stage 1 are sorted into a usable form. This sorting is used in such a way that instead of comparing each point (r_1, n_1, r_2, n_2) produced in Stage 3 with the whole set of the exceptional points produced in Stage 1, we just compare that specific point (r_1, n_1, r_2, n_2) with that subset of exceptional points which has the same value of r_1 in the first place (first dimension of the four dimensional integer space).

Stage 3:

The simulation procedures have been carried out in this stage using the same set of exceptional points produced in Stage 2 and the same value of δ_0 used to produce that set. If the point produced through this stage is coincident with one of the exceptional points then a different decision, depending on the type of the optimal point (that is whether it is S_1, S_2, B_1 or B_2 in the case of OPT_1 or S_1, S_2 or B in the case of OPT_2), from that which used with suboptimal scheme should be taken. If the point does not exist in the list of the exceptional points, that means we follow the suboptimal scheme.

Note: In the OFSS we follow the same stages, but only S_1 and S_2 are involved in the comparison. It is noted that the number of exceptional points in OFSS is zero for N = 10(10)50 for all examples we considered. This means that the suboptimal fixed sample size scheme based on the posterior estimates and the optimal fixed sample size scheme OFSS have the same performance. This can be explained by the limiting behaviour of the posterior estimates $\frac{r_i}{n_i}$ of p_i $(i = 1, 2)$; since the procedure takes all N observations and in the limit $\frac{r_i}{n_i} \to p_i$.

6.3 Results and discussion

In this section we compare and discuss the MC estimates of the performance characteristics, given in section 6.1, of the Binomial selection schemes OPT_1, OPT_2, OFSS to give a comprehensive picture of the performance of these schemes.

The investigations have been carried out for generated p_1, p_2, fixed p_1, p_2 with $p_{[2]} - p_{[1]} = 0.2$ and for different values of N. Since the performance characteristics P(CS), E(M) etc. depend on the parameters of interest p_1, p_2 and the particular selection procedure under investigation, there are some complexities associated with these comparisons. For example, if OPT_1 and OPT_2 have similar P(CS) for particular values of p_1 and p_2, then they can be easily compared in terms of their E(M) or $E(N_{(1)})$. Also if they achieve similar E(M) or $E(N_{(1)})$ then they can be compared in terms of their P(CS). Unfortunately the situation is more complicated than described above. Suppose both schemes are different in P(CS) and E(M) then it is not possible to say that one of them is better than the other. In these situations the preference of one on the other can only be stated when one is interested in one of these performance characteristics. However, in comparing procedures which do not have same P(CS) or E(M), the measures $E(N^*_{(1)})$ and $E(R^*)$ give more direct comparison since they are functions of P(CS), E(M), $E(N_{(1)})$ and calculated as if sampling continues up to N. In addition, indices such as P(CS)/E(M), $E(N_{(1)})/E(M)$, E(R)/E(M), $E(N^*_{(1)})/N$ and $E(R^*)/N$ might also be helpful in comparing these procedures.

6.3.1 The MC estimates of P(CS)

The probability of correct selection increases with increasing N for all schemes for all sets of simulations. The scheme OFSS is the best in all cases but this is to be expected since it uses all N observations. The values for generated p_1 and p_2 in Table (6.1) for schemes OPT_1 and OPT_2 (n = 1) are broadly similar, while OPT_2 (n = 1) is better than OPT_1 for small N and for fixed p_1 and p_2 in Table (6.3), this may be a reflection of the larger expected sample sizes for OPT_2 (n = 1) in these cases.

It was noted in Bechhofer and Kulkarni (1982) that for all N the P(CS) for ($p_{[1]} = h$, $p_{[2]} = h + \Delta'$) equals P(CS) for the symmetric values of $p_{[1]}$ and $p_{[2]}$ ($p_{[1]} = 1 - h - \Delta'$, $p_{[2]} = 1 - h$) for all (h, Δ') with $h > 0$, $\Delta' > 0$ and $h + \Delta' < 1$. We confirm that this property holds for our results in Table (6.3) where $\Delta' = 0.2$ and h = (0.1)(0.2)0.7. For example P(CS) for (.3, .5) is similar to P(CS) for (.5, .7) for all schemes and all N. The slight variations in P(CS) may be due to the simulation fluctuations.

Finally, Table (6.2) demonstrates the effect of grouping on the performance characteristics of OPT_2. It is clear from the table that there is little increase in P(CS) as the group size 2n increases.

Table (6.1)

Performance characteristics for OPT_1, OPT_2 (n = 1) and OFSS, where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear loss function, uniform priors with generated p_1 and p_2 for values of truncated sample size N.

performance characteristics	N	scheme OPT_1	scheme OPT_2 (n = 1)	scheme OFSS
P(CS)	10	.8169	.8168	.8155
	20	.8672	.8574	.8624
	30	.8727	.8724	.8886
	40	.8833	.8771	.9019
	50	.8870	.8810	.9072
E(M)	10	6.8233	7.1914	
	20	9.9489	10.1992	
	30	11.5234	11.3934	
	40	12.4989	11.9162	
	50	12.7617	12.0538	
$E(N_{(1)})$	10	3.4083	3.5957	
	20	4.9852	5.0996	
	30	5.7478	5.6967	
	40	6.2501	5.9581	
	50	6.3870	6.0269	
E(R)	10	3.4427	3.5940	4.9950
	20	4.9674	5.1259	9.9173
	30	5.7900	5.6219	15.0641
	40	6.2570	5.9458	20.0035
	50	6.2600	5.9897	24.9950

Table (6.1) continued

Performance characteristics for OPT_1, OPT_2 (n = 1) and OFSS, where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear loss function, uniform priors with generated p_1 and p_2 for values of truncated sample size N.

performance characteristics	N	scheme		
		OPT_1	OPT_2 (n = 1)	OFSS
$E(N^*_{(1)})$	10	4.1670	4.0985	
	20	6.0783	6.5048	
	30	7.8657	7.7543	
	40	9.3248	8.7907	
	50	10.3832	9.9993	
$E(R^*)$	10	5.8265	5.7081	
	20	11.6874	12.5290	
	30	18.5740	18.2901	
	40	25.0897	24.3173	
	50	31.2290	30.8499	

Table (6.2)

Performance characteristics of OPT_2 with different group size 2n, N = 120, $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear loss function, uniform priors and generated p_1, p_2.

2n	Performance characteristics					
	P(CS)	E(M)	$E(N_{(1)})$	E(R)	$E(N^*_{(1)})$	$E(R^*)$
2	.8804	12.1368	6.0684	6.0510	18.8836	77.4619
4	.8764	13.3496	6.6748	6.6258	19.8228	77.3455
6	.8836	14.5890	7.2945	7.2849	19.5435	77.6682
10	.8895	17.0700	8.5350	8.4847	20.0510	77.6516
12	.8851	17.4888	8.7444	8.7589	20.5980	77.1777
20	.8894	23.7180	11.8590	11.7778	22.7210	75.5389
24	.8917	25.6536	12.8268	12.8294	23.1060	74.8608
30	.8983	31.3920	15.6960	15.6958	24.7950	74.2976
40	.9007	40.2720	20.1360	20.1633	28.0640	72.1687
60	.9225	60.0000	30.0000	29.8182	34.6500	68.7446
120	.9409	120.0000	60.0000	60.0182	60.0000	60.0182

Table (6.3)

P(CS) for OPT_1, OPT_2 (n = 1) and OFSS where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear loss function, p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = 0.2$ starting from $p_{[2]} = 0.3$ with uniform priors for values of truncated sample size N.

scheme	N	$p_{[2]} - p_{[1]} = 0.2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
OPT_1	10	.8207	.7293	.7353	.7790
	20	.8701	.8007	.8118	.8665
	30	.8890	.8419	.8474	.8929
	40	.9024	.8528	.8562	.9032
	50	.9006	.8691	.8619	.9050
OPT_2 (n = 1)	10	.8748	.7964	.7940	.8756
	20	.8934	.8224	.8271	.8964
	30	.9064	.8341	.8366	.9088
	40	.9093	.8421	.8443	.9098
	50	.9125	.8441	.8465	.9137
OFSS	10	.9015	.8369	.8406	.9005
	20	.9294	.8756	.8804	.9305
	30	.9511	.9067	.9123	.9527
	40	.9679	.9292	.9314	.9709
	50	.9808	.9455	.9468	.9789

6.3.2 The MC estimates of E(M)

In what follows, we shall judge the relative merits of the procedures OPT_1, OPT_2 (n = 1) and OFSS on the basis of the expected sample sizes E(M).

In Tables (6.1) and (6.4), the expected sample sizes for the sequential schemes indicate substantial savings in observations over OFSS especially for large N. The expectation increases less rapidly as N increases this is further evidence for the use of truncation since a relatively small E(M) suggests that there is a small probability of reaching the truncation point. As expected under priors different from uniform we gained some reduction in E(M). For example, using OPT_1 with $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear losses with N = 50, E(M) decreases from 14.0639 under (Be(1, 2) v Be(1, 2)) priors to 9.9321 under (Be(1, 3) v Be(1, 2)) priors for ($p_{[1]} = .7$, $p_{[2]} = .9$).

From the results of E(M) in Table (6.4), the percent savings, defined by [N - E(M)]100/N, with fixed ($p_{[1]} = .1$, $p_{[2]} = .3$) increases from 24.15% to 70% under OPT_1 and from 19.54% to 68.97% under OPT_2 (n = 1) as N increases from 10 to 50.

Obviously, in OPT_2 (n = 1) as the group size 2n increases, E(M) increases and the increase will be considerable for large values of 2n as given in Table (6.2).

6.3.3 The MC estimates of $E(N_{(1)})$ and $E(N^*_{(1)})$

$E(N_{(1)})$ is important in clinical trials where p_i denotes the probability of a cure using treatment i (i = 1, 2). In this application the primary goal is to minimize the use of poorer treatment and even slight decreases in $E(N_{(1)})$ are important.

Tables (6.1) and (6.5) show that the values of $E(N_{(1)})$ in OPT_1 and OPT_2 ($n = 1$) are small relative to N, particularly for large values of N. In order to compare $E(N_{(1)})$ for OPT_1 with the other two schemes, it is helpful to calculate $E(N_{(1)})/E(M)$ which is close to ½ for generated p_1 and p_2 indicating that it is close to OPT_2 ($n = 1$), but it displays a considerable variations for fixed p_1 and p_2 always greater than ½ for (.1, .3) and about ½ for (.3, .5) and less than ½ for large N for other two pairs. However, $E(N_{(1)})/E(M)$ has roughly constant values for all N under the scheme OPT_1 for each particular p_1 and p_2.

The values of $E(N^*_{(1)})$ are given in Tables (6.1) and (6.6). For the scheme OFSS, the value of $E(N_{(1)})$ is N/2.

It is noted that $E(N^*_{(1)})/N$ decreases as N increases and is less than ½ for both OPT_1 and OPT_2 ($n = 1$) for all pairs of values of p_1 and p_2.

6.3.4 The MC estimates of E(R) and E(R*)

From Table (6.7) we note that for the schemes OPT_1, OPT_2 ($n = 1$) and OFSS, E(R) increases as $p_{[1]}$ increases ($p_{[2]} - p_{[1]} = 0.2$) for all values of N. E(R)/E(M) increases rapidly as $p_{[1]}$ increases ($p_{[2]} - p_{[1]} = 0.2$) for all N and for all schemes. In addition, the ratio has approximately the same value for all N for each particular scheme and for each pair of values of p_1 and p_2. This can also be inferred from the equation (6.1.7) from which

$$\frac{E(R)}{E(M)} = p_{[2]} - \frac{E(N_{(1)})}{E(M)}(p_{[2]} - p_{[1]}) \qquad (6.3.1)$$

hence if $E(N_{(1)})/E(M)$ is constant, then $E(R)/E(M)$ is also constant for fixed p_1 and p_2. $E(R^*)$ is equal to N/2 under the scheme OFSS with p_1 and p_2 are generated.

The ratio $E(R^*)/N$ has similar behaviour to $E(R)/E(M)$ with little increase in its values.

It is of interest to compare the value of $E(R^*)$ with optimal value using dynamic programming equations (Jones (1975)). From Table (6.1), for N = 50 we have $E(R^*) = 31.2290$ while the optimal value using dynamic programming equations is 31.9967.

Table (6.4)

E(M) for OPT_1, OPT_2 (n = 1), where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear loss function, p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = 0.2$ starting from $p_{[2]} = 0.3$ with uniform priors for values of truncated sample size N.

scheme	N	$p_{[2]} - p_{[1]} = 0.2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
OPT_1	10	7.5853	7.2644	7.2616	7.4731
	20	11.8627	11.6669	11.5665	11.7388
	30	12.6749	14.6368	14.7295	12.7066
	40	14.1935	16.0715	16.1046	13.9989
	50	14.3071	16.5635	16.4794	14.0639
OPT_2 (n = 1)	10	8.0460	7.6046	7.5960	8.0736
	20	12.4988	11.4586	11.4776	12.5880
	30	14.5388	12.9194	12.8834	14.6262
	40	15.1440	13.5332	13.4584	15.2148
	50	15.5150	13.7446	13.6720	15.5360

Table (6.5)

$E(N_{(1)})$ for OPT_1, OPT_2 ($n = 1$), where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear loss function, p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = 0.2$ starting from $p_{[2]} = 0.3$, uniform priors for values of truncated sample size N.

scheme	N	$p_{[2]} - p_{[1]} = 0.2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
OPT_1	10	4.8225	3.9953	3.7672	3.8407
	20	7.1313	6.1742	5.6763	5.3138
	30	7.5971	7.7018	7.0864	5.4477
	40	8.2961	8.3887	7.7689	6.1039
	50	8.3666	8.6915	7.9637	6.0998
OPT_2 ($n = 1$)	10	4.0230	3.8023	3.7890	4.0368
	20	6.2494	5.7293	5.7388	6.2940
	30	7.2694	6.4597	6.4417	7.3131
	40	7.5720	6.7666	6.7292	7.6074
	50	7.7575	6.8723	6.8360	7.7680

Table (6.6)

$E(N^*_{(1)})$ for OPT_1, OPT_2 (n = 1), where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear loss function, p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = 0.2$ starting from $p_{[2]} = 0.3$, uniform priors for values of truncated sample size N.

scheme	N	$p_{[2]} - p_{[1]} = 0.2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
OPT_1	10	5.5322	5.2096	4.8538	4.4217
	20	8.0516	7.7737	7.2075	6.2861
	30	9.5955	10.1734	9.4286	7.4386
	40	10.9113	11.9898	11.3858	8.7388
	50	12.0789	13.2252	12.7472	9.6314
OPT_2 (n = 1)	10	4.4822	4.6205	4.6158	4.4872
	20	7.4450	7.8015	7.8144	7.4450
	30	8.9020	9.4095	9.3547	8.8877
	40	9.6020	10.8346	10.7616	9.5706
	50	10.5339	12.4571	12.3394	10.5012

Table (6.7)

E(R) for OPT_1, OPT_2 (n = 1), where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear loss function, p_1 and P_2 are fixed with $p_{[2]} - p_{[1]} = 0.2$ starting from $p_{[2]} = 0.3$ with uniform priors for values of truncated sample size N.

scheme	N	$p_{[2]} - p_{[1]} = 0.2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
OPT_1	10	1.3279	2.8589	4.3429	5.9678
	20	2.1383	4.6202	6.9722	9.5234
	30	2.2807	5.7830	8.9016	10.3639
	40	2.5958	6.3385	9.7180	11.3900
	50	2.6086	6.5389	9.9436	11.4441
OPT_2 (n = 1)	10	1.6209	3.0793	4.5686	6.4617
	20	2.4995	4.6048	6.8998	10.0789
	30	2.9026	5.1886	7.7458	11.7002
	40	3.0179	5.4337	8.0894	12.1648
	50	3.0923	5.5180	8.2196	12.4218
OFSS	10	2.0026	4.0251	6.0192	8.0197
	20	3.9851	7.9787	11.9941	15.9890
	30	5.9760	12.0026	18.0207	24.0005
	40	7.9997	15.9825	24.0099	32.0067
	50	9.9909	20.0108	30.0354	40.0442

Table (6.8)

$E(R^*)$ for OPT_1, OPT_2 ($n = 1$), where $K_1 = K_2 = 1000$, $C_1 = C_2 = 1$, linear loss function, p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = 0.2$ starting from $p_{[2]} = 0.3$ with uniform priors for values of truncated sample size N.

scheme	N	$p_{[2]} - p_{[1]} = 0.2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
OPT_1	10	2.2465	4.6111	6.8485	8.4124
	20	4.2793	8.6793	12.7271	16.6409
	30	7.0985	13.2195	19.6036	26.3361
	40	9.9316	17.9616	26.4602	35.2289
	50	12.7374	22.7929	33.1399	43.8339
OPT_2 ($n = 1$)	10	2.1388	4.6165	6.6356	8.7484
	20	4.9743	9.7336	14.0930	18.6461
	30	7.3479	13.4973	19.4260	26.3547
	40	9.6379	17.7867	25.5196	33.4864
	50	12.4378	22.4624	32.2057	42.2934

6.4 Conclusion

On the basis of the results obtained from this study we can draw some general conclusions about the performance of the schemes.

We noted that the sequential schemes OPT_1 and OPT_2 have remarkable performance. They perform considerably better than OFSS. Very substantial savings in sample size can be gained if they are used instead of OFSS scheme. These savings increases as N or $(p_{[1]}, p_{[2]})$ or both increase. Moreover, the small values of $E(N_{(1)})$ and $E(N^*_{(1)})$ provides a numerical evidence about the usefulness of OPT_1 and OPT_2 in clinical trials where it is desirable to minimize the expected number of patients subjected to the inferior treatment. For generated probabilities, OPT_1 is generally better than OPT_2 $(n = 1)$, while for fixed probabilities OPT_2 $(n = 1)$ is better scheme except for large probabilities.

The scheme OPT_1 has desirable $E(N_{(1)})$ behaviour relative to OPT_2 but less desirable P(CS) behaviour. It appears that the decrease in E(M) for OPT_1 is purchased at the cost of a decrease in P(CS). The little increase in P(CS) in OPT_2 as the group size 2n increases suggests that we can gain a considerable reduction in E(M) and $E(N_{(1)})$ if we use smaller group sizes as shown in Table (6.2).

Based on the values of $E(N_{(1)})/E(M)$ and $E(N^*_{(1)})/N$ we can conclude that OPT_1 performs better for larger values of $p_{[1]}$ and $p_{[2]}$ and OPT_2 $(n = 1)$ for smaller values of $p_{[1]}$ and $p_{[2]}$. Naturally these schemes require larger E(M) as the values of $p_{[1]}$ and $p_{[2]}$ get closer.

Broadly speaking, the choice of selection scheme depends on the objectives of the experimenter. If the correct decision without regard to the cost of sampling, the outcome of the observation, the population, then P(CS) is the most important criterion to be chosen to judge the appropriate selection scheme. Among our schemes, we should choose OFSS for this purpose as it achieves the best P(CS); but this must be balanced against using all N observations with consequent worsening of the other criteria including risk. If the cost of sampling is the most important factor in the experimentation irrespective of the outcome of observation or the population we sample from, then $E(M)$ is the ideal criterion for selecting the best scheme. If the outcome of the observation is of interest as in the clinical trials where the ethical considerations should be taken into account, then $E(N_{(1)})$ and $E(N^*_{(1)})$ would be preferred.

It appears from the results and discussion that OPT_2, with small and moderate values of n, compared well with OPT_1 in most of the performance characteristics; further, OPT_2 is easier to implement in practical situations, this suggests that OPT_2 should be recommended.

CHAPTER 7

SIMULATION STUDIES OF SOME BAYESIAN SEQUENTIAL SUBOPTIMAL SCHEMES FOR CHOOSING THE BETTER OF TWO BINOMIAL POPULATIONS

7.0 Introduction and summary

This chapter has two aims. The first is to study the performance of the suboptimal schemes which are based on the posterior estimates of the unknown probabilities of the populations, given in chapter 5, under the criteria (a - f) in section (6.2).

The second aim is to make comparisons between these suboptimal schemes and some alternative schemes given in the literature, namely those proposed by Bechhofer and Kulkarni (1981). The results of a large scale Monte Carlo simulation are presented.

Some modifications to the stopping rule of Bechhofer and Kulkarni (1981) are suggested and several selection schemes arising from using different sampling rules, stopping rules and terminal decision rules are investigated. Since extensive numerical results are involved, the discussion is only limited to the uniform prior case.

Deciding which is the best design comes down to the problem of balancing the two most important criteria (a), (b); usually, but not always P(CS) increases with E(M) so an index such as P(CS)/E(M) could be calculated giving an idea of the information per observation for each design. (c) and (d) are related to (b) and usually increase with E(M), so again indices such as $E(N_{(1)})/E(M)$ and $E(R)/E(M)$ may give some indication of the best design; (e) and (f) are related to both (a), (b) and give a way of comparing all the designs over the full N observations, they also give an idea of how the designs perform if used as the first stage of a two stages design where the chosen population is sampled for all the remaining observations at the second stage.

As in chapter 6, the values of the performance measures were calculated from the results of Monte Carlo simulations for 10,000 trials in two cases, for generated values of p_1, p_2 from a prior distribution (uniform in the cases considered) and for fixed values of p_1, p_2.

The contents of this chapter can be summarized as follows.

The performance of the suboptimal selection δ-schemes is studied in section 7.1.

In section 7.2 the selection schemes of Bechhofer and Kulkarni (1981) are presented and some modifications to their stopping rule are discussed.

Section 7.3 discusses some further selection schemes using various sampling and stopping rules.

Some comparisons and discussion of various schemes, considered in the previous sections, are given in section 7.4.

7.1 Performance characteristics of the schemes δ_1, δ_G and FSS

In this section, attention is confined to the study of suboptimal schemes based on the posterior expectation of p_i ($i = 1, 2$) using the Monte Carlo simulation technique. The purely sequential schemes δ_1, which are constructed from the sampling rules A_1, A_2, A_3, A_4, A_5, A_6 in conjunction with the stopping rule DS and the terminal decision rule DT, the group sequential schemes δ_G and the fixed suboptimal sample size FSS are considered. As we mentioned, in the purely sequential case the sampling scheme is a design since a choice has to be made between sampling the two populations, whilst in the group sequential scheme this becomes a problem in optimal stopping for $n = 1(1)(N/2 - 1)$ and for $n = N/2$ it is a fixed sample size scheme. To assess the properties of these schemes we need to study the effect of the parameter δ_0 on the behaviour of the schemes. The performance measures P(CS), E(M), $E(N_{(1)})$ and E(R) are increasing functions of δ_0 but the rate of increase is large for small values of δ_0 and small for large values of δ_0. Therefore, sometimes, an increase in the value of δ_0 has little effect on P(CS), but allowing a small decrease in P(CS) can be compensated for by a large reduction in E(M) and $E(N_{(1)})$ as it is clear from Table (7.1).

Furthermore, $E(N^*_{(1)})$ increases as δ_0 increases under δ_1 (A_1), δ_1 (A_2) and δ_1 (A_3) while it decreases as δ_0 increases under δ_1 (A_4), δ_1 (A_5) and δ_1 (A_6). On the contrary, $E(R^*)$ decreases as δ_0 increases under δ_1 (A_1), δ_1 (A_2) and δ_1 (A_3) while it increases as δ_0 increases under δ_1 (A_4), δ_1 (A_5) and δ_1 (A_6). Generally speaking, there is very little change in the values of each performance characteristic for very large values of δ_0 for all schemes.

Graphically, the effect of δ_0 on the performance measures of the schemes δ_1 (A_1) and δ_1 (A_4) are displayed in Figures (7.1 - 7.6).

Table (7.1)

Performance characteristics of the schemes δ_1 (A_1) and δ_1 (A_4) for N = 50, different values of δ_0, generated p_1 and p_2 under uniform priors.

δ_0	P(CS)	E(M)	$E(N_{(1)})$	E(R)	$E(N^*_{(1)})$	$E(R^*)$
	δ_1 (A_1)					
0.2	.7915	6.9658	4.3673	4.3950	13.2742	30.1048
0.3	.8200	16.1731	11.0251	9.0646	16.4300	29.3295
0.4	.8404	30.8971	22.6970	14.3464	23.3590	26.0296
0.5	.8386	43.9477	34.1234	17.4011	34.3270	21.4271
0.6	.8439	48.3951	38.6985	18.0803	38.7180	19.1608
0.7	.8377	49.9954	39.9608	18.3348	39.9608	18.3378
0.8	.8426	50.0000	40.2920	18.1817	40.2920	18.1817
0.9	.8453	50.0000	40.3970	18.1470	40.3970	18.1470
	δ_1 (A_4)					
0.2	.7855	7.1318	2.6976	2.6397	11.7609	30.5980
0.3	.8234	17.5924	5.5571	7.9679	10.4297	30.9812
0.4	.8350	30.9275	8.3630	16.6649	10.0362	31.5009
0.5	.8383	43.2221	9.6776	26.0194	9.9837	31.5083
0.6	.8409	48.4891	9.7370	30.3781	9.7581	31.7580
0.7	.8423	49.9956	9.8312	31.7556	9.8312	31.7593
0.8	.8404	50.0000	9.9213	31.6350	9.9213	31.6350
0.9	.8401	50.0000	9.8210	31.7262	9.8210	31.7262

NB: If uniform priors are used then $\left|\frac{r_1}{n_1} - \frac{r_2}{n_2}\right| = \frac{1}{6}$ after 1 observation and for $\delta_0 < \frac{1}{6}$ only 1 observation is taken; hence we start at $\delta_0 = .2$.

We conclude that it is necessary to find a compromise between δ_0 which gives earlier stopping resulting in smaller P(CS) and larger δ_0 giving later stopping with larger P(CS). In our particular examples, where p_1, p_2 are generated from uniform distribution or fixed with $p_{[2]} = p_{[1]} + .2$ ($p_{[1]} = .1, .3, .5, .7$) with uniform priors, the value $\delta_0 = .4$ is a good compromise if we wish to judge the overall performance of the schemes under all measures we considered.

The comparison between the sampling rules ($A_1 - A_6$) revealed little differences for P(CS) and E(M) but large differences between the group ($A_1 - A_3$) and ($A_4 - A_6$) under the criteria $E(N_{(1)})$ and E(R) where p_1, p_2 are generated as given in Table (7.2). However, the differences between the two groups are more pronounced when the values of p_1, p_2 are fixed as shown in Table (7.3). Both tables show that the performance measures are increasing functions of N. As a group the sampling rules ($A_4 - A_6$) have the property of sampling from the population which has the greater posterior expected success probability and hence the property of stay on the winner. The group of the sampling rules ($A_1 - A_3$) has the property of sampling from the population which has the smaller posterior expected success probability and hence it has the property of stay on the loser. In general, the sampling rule A_2 performs better than other sampling rules within the group ($A_1 - A_3$) and A_5 performs better than other sampling rules within the group ($A_2 - A_6$). Therefore these two sampling rules are chosen as the candidates representing these two groups for future comparisons.

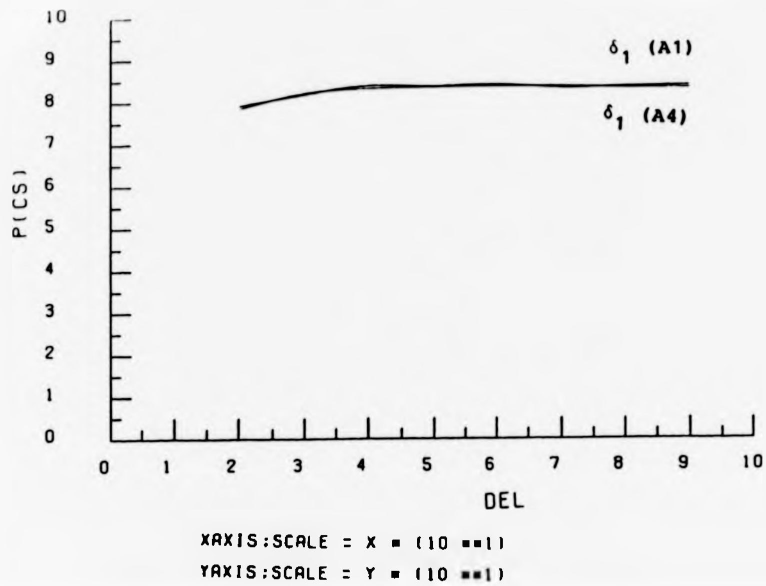

Fig. 7.1 P (CS) as a function of DEL (δ_0) for the schemes δ_1 (A_1) and δ_1 (A_4) when N = 50 with uniform priors on p_1 and p_2.

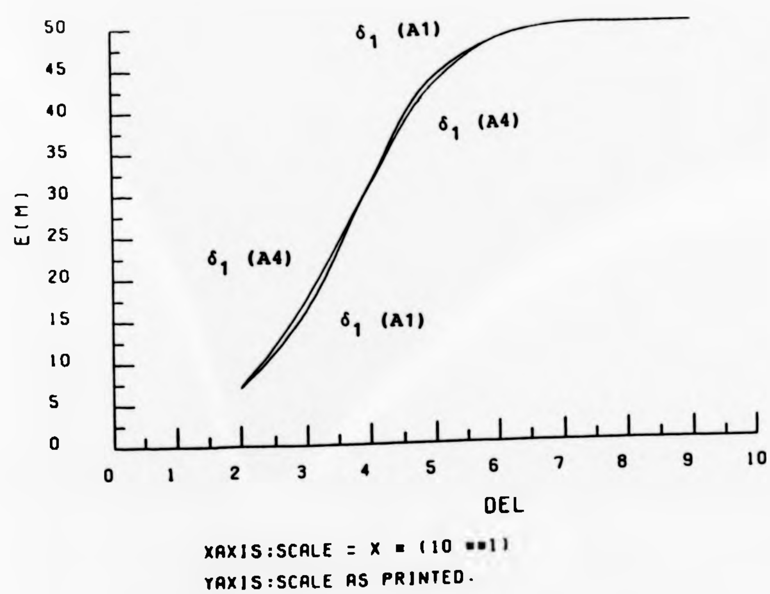

Fig. 7.2 E(M) as a function of DEL (δ_0) for the schemes δ_1 (A_1) and δ_1 (A_4) when N = 50 with uniform priors on p_1 and p_2.

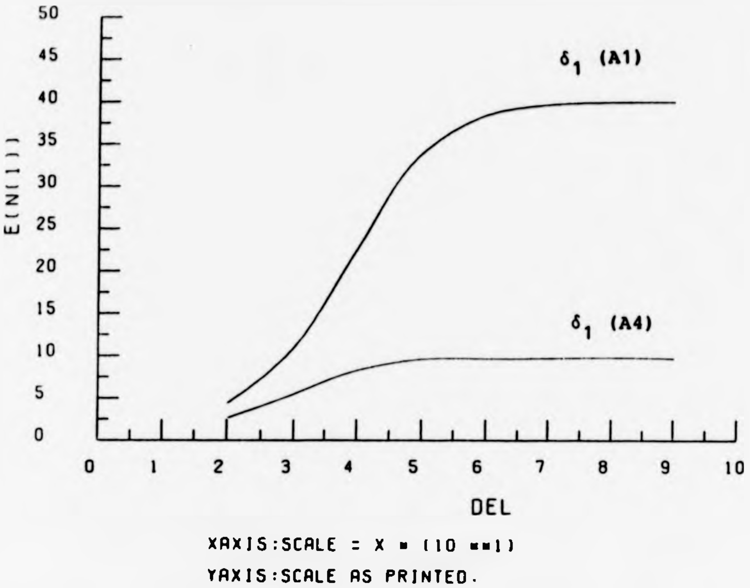

Fig. 7.3 $E(N_{(1)})$ as a function of DEL (δ_0) for the schemes δ_1 (A_1) and δ_1 (A_4) when N = 50 with uniform priors on p_1 and p_2.

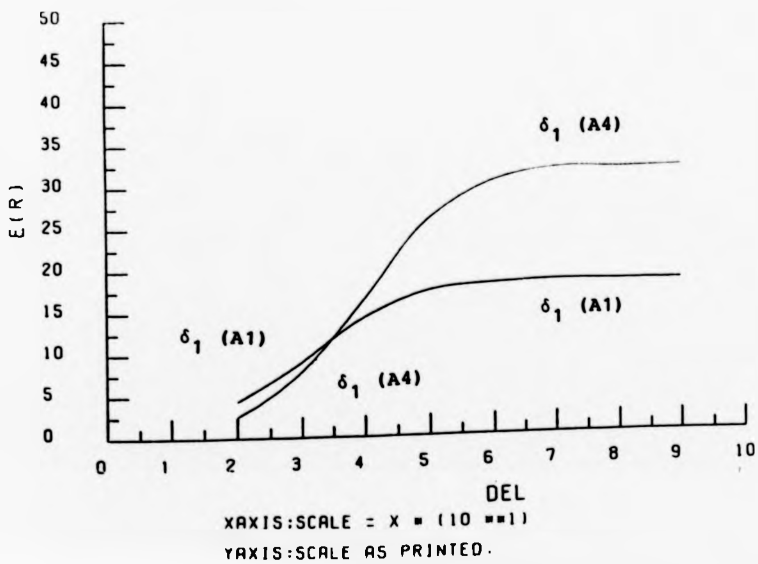

Fig. 7.4 E(R) as a function of DEL (δ_0) for the schemes δ_1 (A_1) and δ_1 (A_4) where N = 50 with uniform priors on p_1 and p_2.

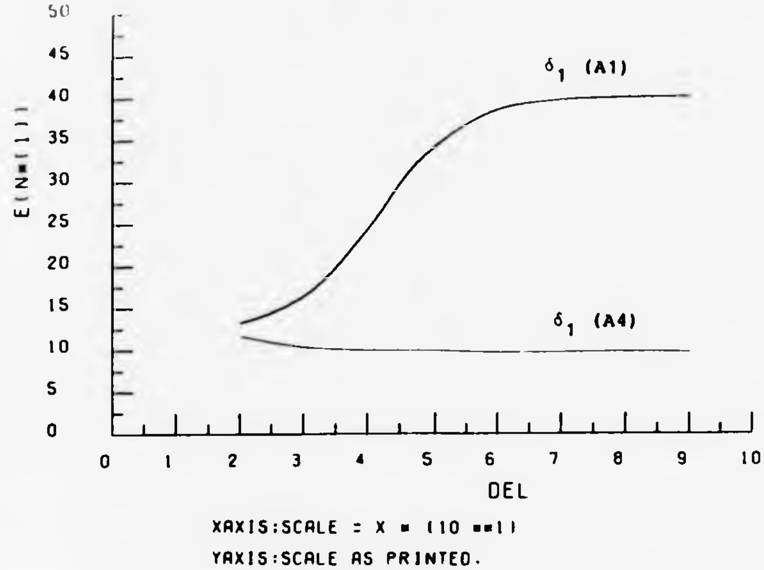

Fig. 7.5 $E(N^*_{(1)})$ as a function of DEL (δ_0) for the schemes δ_1 (A_1) and δ_1 (A_4) where N = 50 with uniform priors on p_1 and p_2.

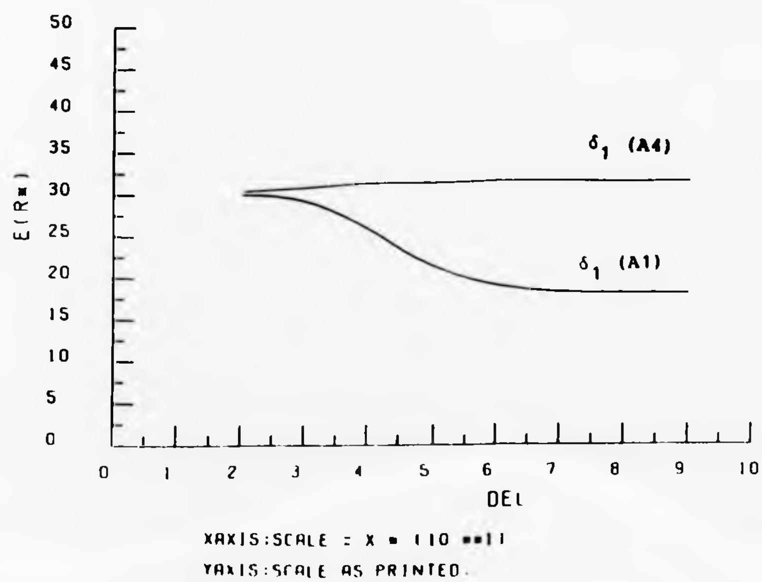

Fig. 7.6 $E(R^*)$ as a function of DEL (δ_0) for the schemes δ_1 (A_1) and δ_1 (A_4) where N = 50 with uniform priors on p_1 and p_2.

Table (7.2)

Performance characteristics of the schemes δ_1 (A_1), δ_1 (A_2), δ_1 (A_3), δ_1 (A_4), δ_1 (A_5) and δ_1 (A_6), N = 10(10)50, generated p_1 and p_2, δ_0 = .4 under uniform priors.

N	P(CS)	E(M)	$E(N_{(1)})$	E(R)	$E(N^*_{(1)})$	$E(R^*)$
	δ_1 (A_1)					
10	.7783	8.4175	5.6727	3.6787	5.8483	4.5110
20	.8148	14.3425	10.0710	6.3240	10.5926	9.6526
30	.8278	20.0799	14.4457	9.0880	15.4304	15.0340
40	.8350	25.5973	18.5549	11.7834	19.8803	20.4894
50	.8404	30.8971	22.6970	14.3464	24.3590	26.0296
	δ_1 (A_2)					
10	.7985	8.3841	5.7580	3.5980	5.9098	4.4691
20	.8220	14.3482	10.1881	6.3707	10.7044	9.6853
30	.8369	19.8040	14.2042	8.9483	15.0304	15.0423
40	.8337	25.2980	18.3685	11.6569	19.7348	20.4243
50	.8340	30.6802	22.5224	14.4248	24.3626	26.0657
	δ_1 (A_3)					
10	.7859	8.3738	5.6862	3.6284	5.8560	4.4850
20	.8026	14.5078	9.9910	6.4267	10.4965	9.6514
30	.8285	20.0307	14.3025	9.0289	15.1380	15.0316
40	.8303	25.6234	18.3823	11.7619	19.6894	20.5047
50	.8242	31.0301	22.0204	14.6221	23.7343	26.3505

Table (7.2) continued

Performance characteristics of the schemes δ_1 (A_1), δ_1 (A_2), δ_1 (A_3), δ_1 (A_4), δ_1 (A_5) and δ_1 (A_6), N = 10(10)50, generated p_1 and p_2, δ_0 = .4 under uniform priors.

N	P(CS)	E(M)	$E(N_{(1)})$	E(R)	$E(N^*_{(1)})$	$E(R^*)$
	δ_1 (A_4)					
10	.7871	8.5485	2.7453	4.9222	2.8836	5.7948
20	.8116	14.7897	4.4247	8.2698	4.8960	12.1146
30	.8310	20.4250	5.6631	11.2770	6.4250	18.5783
40	.8293	26.0010	7.0613	14.0293	8.4593	24.8011
50	.8350	30.9275	8.3630	16.6649	10.0362	31.5009
	δ_1 (A_5)					
10	.7995	8.5690	2.6549	4.9502	2.7957	5.8199
20	.8203	14.7466	4.2827	8.3177	4.7798	12.1995
30	.8323	20.6166	5.6993	11.4063	6.5295	18.5435
40	.8392	25.8771	7.0313	14.1621	8.1252	25.1099
50	.8422	31.1810	8.1225	17.0377	9.7048	31.6305
	δ_1 (A_6)					
10	.7792	8.5194	2.8022	4.8416	2.9407	5.7434
20	.8129	14.7640	4.4331	8.2445	4.8953	12.1115
30	.8293	20.5729	5.8811	11.2201	6.6463	18.3901
40	.8320	25.9913	7.2502	14.1251	8.4187	24.9303
50	.8298	31.3757	8.6757	16.8719	19.3609	31.2532

Table (7.3)

Performance characteristics of the schemes δ_1 (A_1), δ_1 (A_2), δ_1 (A_3), δ_1 (A_4), δ_1 (A_5) and δ_1 (A_6), N = 10(10)50, fixed p_1 and p_2 $(p_{[1]} = .1,\ p_{[2]} = .3)$, $\delta_0 = .4$ under uniform priors.

N	P(CS)	E(M)	$E(N_{(1)})$	E(R)	$E(N^*_{(1)})$	$E(R^*)$
	δ_1 (A_1)					
10	.6911	8.3459	5.0699	1.4719	5.3318	1.7223
20	.6932	13.3703	7.6499	2.4845	8.5727	4.0449
30	.6984	17.3503	9.3799	3.3305	10.9676	6.5337
40	.6922	21.4977	10.9469	4.2639	13.2419	9.0740
50	.6996	25.0367	12.4647	5.0167	15.4594	11.6138
	δ_1 (A_2)					
10	.7144	8.3965	5.3201	1.4340	5.5843	1.6657
20	.7205	13.1828	7.9996	2.3493	8.9478	3.9575
30	.7160	17.0649	9.7205	3.1532	11.4586	6.4130
40	.7200	20.9896	11.5383	3.9785	13.8792	8.9353
50	.7126	24.6203	12.8095	4.8294	15.9513	11.5282
	δ_1 (A_3)					
10	.6629	8.2825	4.7589	1.5281	5.0252	1.7917
20	.6590	13.4509	7.1234	2.6119	8.0523	4.1438
30	.6630	17.7984	8.7296	3.5911	10.2298	6.6800
40	.6554	22.0874	10.1434	4.5676	12.4345	9.2041
50	.6549	26.1242	11.3328	5.5707	14.2684	11.8609

Table (7.3) continued

Performance characteristics of the schemes δ_1 (A_1), δ_1 (A_2), δ_1 (A_3), δ_1 (A_4), δ_1 (A_5) and δ_1 (A_6), N = 10(10)50, fixed p_1 and p_2 ($p_{[1]}$ = .1, $p_{[2]}$ = .3), δ_0 = .4 under uniform priors.

N	P(CS)	E(M)	E($N_{(1)}$)	E(R)	E($N^*_{(1)}$)	E(R^*)
	δ_1 (A_4)					
10	.8351	9.6767	3.4733	2.1924	3.4962	2.2747
20	.8852	19.1327	5.3990	4.6530	5.4606	4.8892
30	.8883	28.4906	6.7428	7.1726	6.8397	7.5931
40	.9126	37.9137	7.7298	9.7911	7.8435	10.3821
50	.9380	47.0319	8.1785	12.4372	8.3305	13.2838
	δ_1 (A_5)					
10	.8339	9.6796	3.4983	2.1832	5.5221	2.2644
20	.8846	19.0484	5.4498	4.6130	5.5220	4.8705
30	.8937	28.3949	6.7402	7.1667	6.8281	7.6171
40	.9184	37.7898	7.5509	9.8124	7.6875	10.4350
50	.9361	47.0872	8.2835	12.4450	8.4421	13.2739
	δ_1 (A_6)					
10	.8365	9.6587	3.4379	2.2002	3.4631	2.2870
20	.8860	18.9765	5.3477	4.6238	5.4224	4.9022
30	.8890	28.4164	6.7160	7.1613	6.7940	7.6084
40	.9130	37.7350	7.7130	9.7447	7.8527	10.3828
50	.9340	47.0675	8.3145	12.4146	8.5326	13.2361

Next consider group sequential schemes with n observations on each population at each stage. A series of simulation runs was carried out with $N = 120$ and differing n, it was decided in these cases to use $\delta_0 = .4$ in the sampling/stopping rule. The results are presented in Table (7.4). Varying the group size has little effect on P(CS) but earlier stopping is obtained by decreasing n, in judging the performance one must take into account that the smaller the number of groups the easier and cheaper the scheme would be to implement.

Next we discuss in details the MC estimates of the performance characteristics of the schemes δ_1 (A_2), δ_1 (A_5), δ_2 and FSS.

<u>The MC estimates of P(CS)</u>

For generated p_1 and p_2, P(CS) appears to change very little as N increases under δ_1 (A_2) and δ_1 (A_5) but shows larger differences under δ_2 and FSS as it is clear from Table (7.5).

The index P(CS)/E(M), that is the probability of correct selection per unit observation, can also be used to judge the performance of the schemes. It was found that this ratio is a decreasing function of N; in addition, in terms of this ratio, calculated from Table (7.5), the performance of the schemes can be ordered as follows

$$\delta_2 > \delta_1\ (A_2) > \delta_1\ (A_5) > \text{FSS},$$

where $\delta_2 > \delta_1$ (A_2) means that δ_2 is better than δ_1 (A_2).

As particular examples, we present some results for fixed $p_{[1]}$ and $p_{[2]}$ in Table (7.6), where $p_{[2]} = p_{[1]} + 0.2$,

$p_{[1]}$ = .1, .3, .5, .7. These results show that FSS performs uniformly better than others as far as P(CS) is concerned, but that is compensated for by using larger sample size. Furthermore, it appears from this table that neither δ_1 (A_2) nor δ_1 (A_5) is uniformly better in P(CS) for all p's; rather there is breakeven values for $p_{[1]}$ and $p_{[2]}$ above which δ_1 (A_2) is the better in P(CS) and below which δ_1 (A_5) is the better. This suggests that for $p_{[1]}$, $p_{[2]}$ sufficiently large, δ_1 (A_5) is preferable. Based on the results of P(CS) given in Table (7.6), the performance of the schemes can be ordered as

$$\text{FSS} > \delta_2 > [\delta_1\ (A_2) \simeq \delta_1\ (A_5)].$$

If the performance is measured by the ratio P(CS)/E(M), then the results of Tables (7.6) and (7.7) indicate that the performance has the following ordering

$$\delta_2 > [\delta_1\ (A_2) \simeq \delta_1\ (A_5) \simeq \text{FSS}],$$

where $[\delta_1\ (A_2) \simeq \delta_1\ (A_5)]$ means both schemes have roughly same performance.

<u>The MC estimates of E(M)</u>

For generated p_1, p_2, as can be seen from Table (7.5), the schemes δ_1 (A_2), δ_1 (A_5) and δ_2 give a considerable reduction in E(M). The performance ordering, in terms of E(M), is as follows

$$\delta_2 > \delta_1\ (A_2) > \delta_1\ (A_5) > \text{FSS}.$$

The percent savings in E(M) increases from 16.59% to 38.6%, 14.31% to 37.64%, 18.51% to 45.3% under δ_1 (A_2), δ_1 (A_5), δ_2 respectively as N increases from 10 to 50.

Table (7.7) contains some results of E(M) for fixed p_1 and p_2 which do not favour either scheme uniformly. However, on the basis of these tabulated values one might conclude that δ_1 (A_2) might be better for small values of p_1, p_2 while δ_1 (A_5) is preferable for large values and δ_2 for moderate values. It is intuitively clear that the performance of FSS, as measured by E(M), is poor since E(M) = N.

The MC estimates of $E(N_{(1)})$ and $E(N^*_{(1)})$

The simulation results presented in Table (7.5), where p_1 and p_2 are generated from uniform distribution, indicates that the performance ordering of the schemes is

$$\delta_1\ (A_5) > \delta_2 > \delta_1\ (A_2) > \text{FSS},$$

as measured by $E(N_{(1)})$ and also as measured by $E(N^*_{(1)})$ except for small sample sizes ($N \leq 20$) where FSS > δ_1 (A_2). Moreover from the same table we calculated the ratio $E(N_{(1)})/E(M)$ and found that it is an increasing function of N and its values are greater than ½ under δ_1 (A_2), whilst it is a decreasing function of N and its values are less ½ under the scheme $\delta_1(A_5)$. Therefore the performance of the schemes, based on this ratio with generated p_1 and p_2, can be ordered as follows

$$\delta_1\ (A_5) > [\delta_2 = \text{FSS}] > \delta_1\ (A_2),$$

where $[\delta_2 = \text{FSS}]$ means that δ_2 has exactly the same values of FSS.

From the results in Table (7.5), we can observe that the ratio $E(N^*_{(1)})/N$ decreases as N increases under all schemes and based on this ratio, the performance ordering is

$$\delta_1\ (A_5) > \delta_2 > [\delta_1\ (A_2) \simeq \text{FSS}].$$

The values of $E(N_{(1)})$ and $E(N^*_{(1)})$, for fixed p_1 and p_2 with $p_{[2]} = p_{[1]} + 0.2$ ($p_{[1]} = .1, .3, .5, .7$), are presented in Table (7.8) and (7.9) respectively. Again it can be seen from these tables that δ_1 (A_5) is uniformly superior to the others while δ_2 is better than δ_1 (A_2) and FSS for all $p_{[1]}$ and $p_{[2]}$ except for small $p_{[1]}$ and $p_{[2]}$ ($p_{[1]} = .1$, $p_{[2]} = .3$) and particularly for large sample sizes ($N \geq 20$) where δ_1 (A_2) becomes better than δ_2 and FSS.

Based on the results given in Tables (7.7) and (7.8), the values of $E(N_{(1)})/E(M)$ were calculated and found to be always less than ½ under δ_1 (A_5) and greater than ½ under δ_1 (A_2). Further, under δ_1 (A_5) this ratio is a decreasing function of N for all values of $p_{[1]}$ and $p_{[2]}$ except for very large values of $p_{[1]}$, $p_{[2]}$ ($p_{[1]} \geq .7$, $p_{[2]} = p_{[1]} + 0.2$) where the ratio is an increasing function of N. On the other hand δ_1 (A_2) is a decreasing function of N for all values of $p_{[1]}$, $p_{[2]}$ except for very small values ($p_{[1]} \leq .1$, $p_{[2]} = p_{[1]} + 0.2$).

From Table (7.9), where the values of $E(N^*_{(1)})$ with fixed p_1, p_2 are presented, it appears that $E(N^*_{(1)})/N$ is a decreasing function of N and its values are less than ½ under δ_1 (A_5) and δ_2. However, under δ_1 (A_2) this ratio is also a decreasing function of N for $p_{[1]} \leq .1$ and an increasing function of N for $p_{[1]} > .1$. Under all cases the values of this ratio are greater than ½ except when $p_{[1]}$ and $p_{[2]}$ are small and $N \geqslant 20$.

From the above interpretation we conclude that δ_1 (A_5) is the best among δ_1 (A_2), δ_2 and FSS as measured by the performance characteristics $E(N_{(1)})$ and $E(N^*_{(1)})$ and related indices.

The MC estimates of E(R) and E(R*)

The results in Table (7.5) show clearly that the performance of the schemes, as measured by E(R) is

$$\text{FSS} > \delta_1\ (A_5) > [\delta_2 \simeq \delta_1\ (A_2)].$$

The superiority of FSS here is due to the fact that N observations are taken using FSS. In these situations $E(R^*)$ becomes more important as it gives more direct comparison between the procedures since sampling is continued up to N observations. According to $E(R^*)$ as it is clear from the above table, the performance of the schemes can be ordered as follows

$$\delta_1\ (A_5) > \delta_2 > \delta_1\ (A_2) > \text{FSS}.$$

The results in Table (7.5) show that the ratio E(R)/E(M) is an increasing (a decreasing) function of N and less (more) than ½ under δ_1 (A_2) $(\delta_1\ (A_5))$. The performance of the schemes, based on this ratio with generated p_1, p_2, can be ordered as

$$\delta_1\ (A_5) > [\delta_2 = \text{FSS}] > \delta_1\ (A_2).$$

However, from the same table it can be seen that the ratio $E(R^*)/N$ is an increasing function of N under both δ_1 (A_2) and δ_1 (A_5) and the ordering of the performance is

$$\delta_1\ (A_5) > \delta_2 > FSS > \delta_1\ (A_2).$$

Table (7.10) and (7.11) contain results of E(R) and $E(R^*)$ for fixed p_1 and p_2 with $p_{[2]} = p_{[1]} + 0.2$ ($p_{[1]}$ = 0.1, 0.3, 0.5, 0.7). The results in Table (7.10) show that there is no scheme which is uniformly better for all p_1, p_2 as far as E(R) is concerned. In addition, the results in Table (7.11) show clearly that δ_1 (A_5) is uniformly superior to other schemes as measured by $E(R^*)$ for all p_1, p_2 while δ_2 and FSS are mostly better than δ_1 (A_2).

From the previous discussion we can draw the following conclusions. Although, the group sequential scheme δ_2 is not the best as far as $E(N_{(1)})$ and E(R) are concerned, it has best performance in terms of P(CS); in addition, it has some advantages in its implementation which are not shared by the fully sequential schemes. For example, in fully sequential schemes, the assignment of any observation (patient in the context of clinical trials) to a population (a treatment) depends on the outcome of the previous trial and hence the response must be instantaneous while δ_2 is applicable in situations of delayed response and allows for more observations to be assigned to the populations (for the treatment of several patients) at each stage. On the other hand, δ_1 (A_5) is roughly the best in terms of $E(N^*_{(1)})$ and $E(R^*)$ which are functions of $E(N_{(1)})$, E(M) and P(CS). The balance between the implementation of the scheme and overall optimality of performance suggests that δ_2 should be chosen.

Table (7.4)

Performance characteristics of δ_G for different group size $2n$, $\delta_0 = .4$, generated p_1 and p_2 under uniform priors.

n	Performance characteristics			
	P(CS)	E(M)	$E(N^*_{(1)})$	$E(R^*)$
1	.9303	58.0718	31.6611	73.3140
2	.9360	62.8444	33.0758	73.2526
3	.9377	64.4160	33.8868	73.3240
5	.9383	68.3660	35.1350	72.2385
6	.9406	74.9736	37.7136	71.9849
10	.9442	79.1900	39.7930	70.5314
12	.9405	82.9200	41.5656	69.9170
15	.9426	83.9010	42.0045	69.9070
20	.9412	89.8000	44.9080	68.2223
30	.9460	98.2020	49.1010	66.5262
60	.9433	120.0000	60.0000	59.9400

NB: $E(N_{(1)}) = E(R) = E(M)/2$ since p_1, p_2 are generated from uniform distribution.

Table (7.5)

Performance characteristics of δ_1 (A_2), δ_1 (A_5), δ_2 with δ_0 = .4 and FSS, where N = 10(10)50, generated p_1 and p_2 under uniform priors.

Scheme	N	Performance characteristics					
		P(CS)	E(M)	$E(N_{(1)})$	E(R)	$E(N^*_{(1)})$	$E(R^*)$
δ_1 (A_2)	10	.7985	8.3841	5.7580	3.5980	5.9098	4.4691
	20	.8220	14.3482	10.1881	6.3707	10.7044	9.6853
	30	.8369	19.8040	14.2042	8.9483	15.0304	15.0423
	40	.8334	25.2980	18.3685	11.6569	19.7348	20.4243
	50	.8340	30.6802	22.5224	14.4248	24.3626	26.0657
δ_1 (A_5)	10	.7995	8.5690	2.6549	4.9502	2.7957	5.8199
	20	.8203	14.7466	4.2827	8.3177	4.7798	12.1995
	30	.8323	20.6166	5.6993	11.4063	6.5295	18.5435
	40	.8392	25.8771	7.0313	14.1621	8.1252	25.1099
	50	.8422	31.1810	8.1225	17.0377	9.7048	31.6305
δ_2	10	.8161	8.1494	4.0747	4.0326	4.1705	5.2095
	20	.8581	13.5544	6.7772	6.8256	7.1394	11.3286
	30	.8812	18.4478	9.2239	9.1794	9.8087	17.3839
	40	.8977	23.1030	11.5515	11.4905	12.2981	23.5804
	50	.9036	27.3620	13.6810	13.6224	14.8810	29.8142
FSS	10	.8155					
	20	.8624					
	30	.8886					
	40	.9019					
	50	.9072					

NB: For FSS, $E(N_{(1)}) = E(R) = E(N^*_{(1)}) = E(R^*) = N/2$.

Table (7.6)

P(CS) for δ_1 (A_2), δ_1 (A_5), δ_2 with δ_0 = .4 and FSS, where N = 10(10)50, fixed p_1 and p_2 with $p_{[2]} - p_{[1]}$ = .2 under uniform priors.

Scheme	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
δ_1 (A_2)	10	.7144	.7208	.7431	.8325
	20	.7205	.7449	.7953	.9000
	30	.7160	.7388	.7913	.8921
	40	.7200	.7498	.8058	.9205
	50	.7126	.7576	.8230	.9366
δ_1 (A_5)	10	.8339	.7499	.7285	.7011
	20	.8846	.7811	.7369	.7029
	30	.8937	.7987	.7572	.7179
	40	.9184	.8059	.7537	.7145
	50	.9361	.8248	.7600	.7136
δ_2	10	.8976	.8288	.8289	.8974
	20	.9278	.8635	.8630	.9272
	30	.9500	.8929	.8859	.9466
	40	.9650	.9065	.9073	.9598
	50	.9729	.9212	.9160	.9692
FSS	10	.9015	.8369	.8406	.9005
	20	.9294	.8756	.8804	.9305
	30	.9511	.9067	.9123	.9527
	40	.9679	.9292	.9314	.9709
	50	.9808	.9455	.9468	.9789

Table (7.7)

E(M) for δ_1 (A_2), δ_1 (A_5) and δ_2 with δ_0 = .4, where N = 10(10)50, fixed p_1 and p_2 with $p_{[2]} - p_{[1]}$ = .2 under uniform priors.

Scheme	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
δ_1 (A_2)	10	8.3965	8.6122	9.1000	9.6627
	20	13.1828	15.4604	17.3921	19.1318
	30	17.0649	21.9445	25.4242	28.4841
	40	20.9896	28.1012	33.5759	37.7802
	50	24.6203	33.9322	41.1501	47.1431
δ_1 (A_5)	10	9.6796	9.1091	8.6296	8.6237
	20	19.0484	17.3587	15.5531	13.8699
	30	28.3949	25.2959	22.2797	18.1801
	40	37.7898	33.5240	28.5498	22.4013
	50	47.0872	41.1037	34.7598	26.3338
δ_2	10	9.1526	8.6088	8.5718	9.1170
	20	16.9472	15.1876	15.0588	16.9762
	30	24.2934	21.3102	21.2170	24.5496
	40	31.9248	27.2034	27.0038	31.9636
	50	38.9272	33.0160	32.8198	39.4758

Table (7.8)

$E(N_{(1)})$ for δ_1 (A_2), δ_1 (A_5) and δ_2 with δ_0 = .4, where N = 10(10)50, fixed p_1 and p_2 with $p_{[2]} - p_{[1]}$ = .2 under uniform priors.

Scheme	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
δ_1 (A_2)	10	5.3201	5.3817	5.7261	6.1784
	20	7.9996	10.1113	11.9176	13.7341
	30	9.7205	14.4773	17.8730	21.7517
	40	11.5383	18.7230	24.2427	30.2270
	50	12.8095	22.7181	30.6203	38.8468
δ_1 (A_5)	10	3.4983	3.3518	3.1745	3.2262
	20	5.4498	5.5433	5.4200	5.6304
	30	6.7402	7.4061	7.3112	7.5868
	40	7.5509	9.3779	9.3356	9.7807
	50	8.2835	10.3706	11.1074	12.0704
δ_2	10	4.5763	4.3044	4.2859	4.5585
	20	8.4736	7.5938	7.5294	8.4881
	30	12.1467	10.6551	10.6085	12.2748
	40	15.9624	13.6017	13.5019	15.9818
	50	19.4636	16.5080	16.4099	19.7379

Table (7.9)

$E(N^*_{(1)})$ for δ_1 (A_2), δ_1 (A_5) and δ_2 with δ_0 = .4, where N = 10(10)50, fixed p_1 and p_2 with $p_{[2]} - p_{[1]}$ = .2 under uniform priors.

Scheme	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
δ_1 (A_2)	10	5.5843	5.6828	5.8915	6.2090
	20	8.9478	10.9491	12.3544	13.7896
	30	11.4586	15.8701	18.6387	21.8240
	40	13.8792	20.8311	25.1959	30.3852
	50	15.9513	25.3065	31.9352	39.0712
δ_1 (A_5)	10	5.5221	3.5183	3.4775	3.4531
	20	5.5220	5.9901	6.2805	6.5021
	30	6.8281	8.1722	8.6582	9.0238
	40	7.6875	10.3667	11.3439	11.9868
	50	8.4421	11.6449	13.6864	14.9237
δ_2	10	4.6097	4.4808	4.4853	4.6069
	20	8.5584	8.1122	8.1848	8.6487
	30	12.3019	11.5521	11.6979	12.5260
	40	16.1966	14.9367	14.9635	16.3792
	50	19.7640	18.1264	18.2911	20.2517

Table (7.10)

E(R) for δ_1 (A_2), δ_1 (A_5), δ_2 with δ_0 = .4 and FSS, where N = 10(10)50, fixed p_1 and p_2 with $p_{[2]} - p_{[1]}$ = .2 under uniform priors.

Scheme	N	$p_{[2]} - p_{[1]}$ = .2			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
δ_1 (A_2)	10	1.4340	3.2143	5.2412	7.4595
	20	2.3493	5.7085	9.8090	14.5048
	30	3.1532	8.0942	14.2382	21.2747
	40	3.9785	10.3158	18.6413	27.9347
	50	4.8294	12.4305	22.6713	34.6285
δ_1 (A_5)	10	2.1832	3.8952	5.4047	7.1055
	20	4.6130	7.5749	9.8178	11.3548
	30	7.1667	11.1682	14.1649	14.8541
	40	9.8124	14.8688	18.1467	18.2259
	50	12.4450	18.4576	21.2823	22.1224
δ_2	10	1.8297	3.4482	5.1593	7.3043
	20	3.3932	6.0802	9.0625	13.5957
	30	4.8882	8.5468	12.7545	19.6598
	40	6.3897	10.8994	16.2230	25.5604
	50	7.8061	13.2170	19.7002	31.5619
FSS	10	2.0026	4.0251	6.0192	8.0197
	20	3.9851	7.9787	11.9941	15.9890
	30	5.9760	12.0026	18.0207	24.0005
	40	7.9997	15.9825	24.0099	32.0067
	50	9.9909	20.0108	30.0354	40.0442

Table (7.11)

$E(R^*)$ for δ_1 (A_2), δ_1 (A_5), δ_2 with δ_0 = .4, where N = 10(10)50, fixed p_1 and p_2 with $p_{[2]} - p_{[1]}$ = .2 under uniform priors.

Scheme	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
δ_1 (A_2)	10	1.6657	3.7512	5.7270	7.7327
	20	3.9575	7.6916	11.4247	15.2509
	30	6.4130	11.7256	17.1575	22.6003
	40	8.9353	15.7192	22.8197	29.8768
	50	11.5282	19.8167	28.4673	37.1263
δ_1 (A_5)	10	2.2644	4.2466	6.1343	8.0312
	20	4.8705	8.7388	12.5530	16.3626
	30	7.6171	13.2950	19.0888	24.8298
	40	10.4350	17.8385	25.5394	33.2368
	50	13.2739	22.5755	32.0500	41.6127
δ_2	10	1.9975	4.0865	5.9960	7.9951
	20	4.1970	8.3616	12.2451	16.1549
	30	6.4631	12.6908	18.5330	24.3710
	40	8.6639	17.0087	24.8705	32.5592
	50	10.9617	21.3648	31.1871	40.7764

NB: For FSS, $E(R) = E(R^*)$.

7.2 Bechhofer and Kulkarni (1981) selection schemes with modifications

Bechhofer and Kulkarni (1981) proposed closed sequential schemes for a general class of k-population Binomial selection goals. We confine our attention to the case $k = 2$, that is selecting the better of two Binomial populations. The scheme of Bechhofer and Kulkarni (BK) consists of three rules: a sampling rule (BKR), a stopping rule (BKS) and a terminal decision rule (BKT). The scheme BK takes no more than n^* observation from each population. The reference to this condition is the single-stage procedure of Sobel and Huyett (1957), which takes exactly n^* observations from each population. They (S-H) believed that this choice is reasonable if there is no piror information and it is simple and easy to implement in practical situations.

The single-stage procedure of S-H specifies how to choose n^* to guarantee the indifference zone requirement (P^*, Δ^*-condition) described in section 2.2; while in the scheme BK, the choice of n^* is arbitrary and left to the experimenter.

The scheme BK achieves the same probability of correct selection as does the S-H procedure and it is superior in terms of $E(M)$ to the S-H scheme uniformly in $\underset{\sim}{p} = (p_1, p_2)$.

7.2.1 Description of BK and BK*

In this scheme observations are taken sequentially one at a time. Let N be the total number of observations. When a total of m observations have been taken the experiment is said to be at stage m. Further, let

d_i = number of observations taken from π_i in the first m stages,

c_i = number of successes yielded by π_i in the first m stages,

$s_{i,m}$ = outcome of an observation taken from π_i at stage m if it is a success,

$f_{i,m}$ = outcome of an observation taken from π_i at stage m if it is a failure,

where i = 1, 2 and $0 \leq m \leq N - 1$.

The rules of BK are described as follows:

Sampling rule (BKR): At stage m ($m = d_1 + d_2$, $0 \leq m \leq N - 1$), if the sampling has not stopped then one of the following sampling decisions should be taken.

(a) Take the next observation from π_1 if $d_1 - c_1 < d_2 - c_2$ or ($d_1 - c_1 = d_2 - c_2$ and $c_1 > c_2$) and then proceeds to BKS,

(b) Take the next observation from π_2 if $d_1 - c_1 > d_2 - c_2$ or ($d_1 - c_2 = d_2 - c_2$ and $c_1 < c_2$) and then proceeds to BKS,

(c) If $c_1 = c_2$, select one of the two populations at random and take the next observation from it then proceeds to BKS.

Stopping rule (BKS): Stop sampling at the first stage m at which one of the populations π_1 or π_2 satisfies

$$c_i \geq c_j + n^* - d_j \quad \text{for} \quad i \neq j \ (i, j = 1, 2), \ n^* = N/2. \tag{7.2.1}$$

The left-hand side of the above inequality represents the current number of successes from population π_i while the right-hand side represents the current number of successes from population π_j plus the number of potential successes from π_j if all of the remaining observations $(n^* - d_j)$ from π_j after stage m are successes. In other words, stop sampling as soon as one or both populations have at least as many successes as the maximum possible number of successes at termination from the other population.

Terminal decision rule (BKT): In the inequality (7.2.1),
if $i = 1$ and $j = 2$ make decision D_2,
if $i = 2$ and $j = 1$ make decision D_1,
and in the event of ties, that is $c_1 = c_2$ make decision randomly.

It should be pointed out that the sampling rule BKR is not a play-the-winner sampling rule (PWR). It is a PWR within a cycle, but may not be a PWR as sampling proceeds from one cycle to the next. The following example demonstrates the difference between the two sampling rules.

Example: Let $n^* = 6$ $(N = 12)$, stop if

	π_1	π_2
	$s_{1,1}$	$s_{2,4}$
	$s_{1,2}$	$s_{2,5}$
cycle 1	$f_{1,3}$	$s_{2,6}$
		$f_{2,7}$
cycle 2 is	$f_{1,10}$	$s_{2,8}$
truncated by		$f_{2,9}$
BKS		

According to BKR, the sampling continues with π_2 when the second cycle starts, that is after observation $f_{2,7}$ while in PWR the sampling should be switched to π_1 after that observation.

The conjugate sampling rule BKR* of BKR can be used in conjuction with BKS and BKT rules to obtain another scheme. According to the sampling rule BKR*, we take the next observation from that population with smaller number of successes, breaking ties by using the larger number of failures and randomizing if there is a tie in the number of failures. Let BK* denotes the scheme which consists of the sampling rule BKR*, the stopping rule BKS and the terminal decision rule BKT.

7.2.2 The properties of BK and BK*

Bechhofer and Kulkarni (1981) state several theorems regarding the properties of BK and BK* for the case $k = 2$. These theorems are proved in Kulkarni (1981) and the performance characteristics of these procedures and generalized procedures for $k \geq 3$ are described in Bechhofer and Kulkarni (1982), Bechhofer and Frisardi (1982) and Jennison (1984). Kulkarni and Jennison (1983) strengthened some of these results and obtained generalizations to the case $k \geq 3$.

In this subsection we state these properties of BK and BK* and later we examine some other procedures in the light of some of these properties.

1. $P(CS|BK, p_1, p_2) = P(CS|S\text{-}H, p_1, p_2)$ uniformly in p_1 and p_2.

In fact, Bechhofer and Kulkarni (1981, theorem 5.1) proved that all procedures, which use any sampling rule that takes no more than n^* observations per population in conjunction with BKS and BKT, achieve the same P(CS) as does the single-stage procedure of Sobel and Huyett (1957) which also takes n^* observations per population. Jennison (1983) gave a general theorem covering the above phenomenon and proved that the same P(CS) is attainable for two or more different procedures uniformly in (p_1, p_2).

2. The schemes BK and BK* have the following optimal properties.

Let M be a bounded random variable which is the number of observations taken from the two populations when sampling stops, then among all sequential sampling rules used in conjunction with the stopping rule BKS and the terminal decision rule BKT

(i) BKR minimizes $E(M|p_1, p_2)$ for $p_1 + p_2 \geq 1$, (7.2.2)

(ii) BKR* minimizes $E(M|p_1, p_2)$ for $p_1 + p_2 \leq 1$, (7.2.3)

(iii) BKR minimizes $E(N_{(1)}|p_1, p_2)$ for

$$p_{[2]} \geq \frac{3 - p_{[1]} - \sqrt{(3 - p_{[1]})^2 - 4}}{2}. \qquad (7.2.4)$$

3. The following are additional properties of BK and BK*.

(a) $N/2 \leq M \leq N - 1$ where $n^* = N/2$.

(b) From (a),

$100/N \leq [N - E(M)]100/N \leq 50$ for all p_1, p_2.

Here $[N - E(M)]100/N$ is the percent saving in expected number of observations of BK if used in place of the corresponding single-stage procedure of S-H (1957).

(c) $P\{M = \frac{N}{2} | p_1, p_2\} \to 1$ for $p_{[1]} \to 1$, and

$P\{M = N - 1 | p_1, p_2\} \to 1$ for $p_{[k]} \to 0$.

(d) Populations with small p-values tend to be sampled less frequently.

(e) No special tables of constants are necessary to carry out BK as it is easy to implement.

7.2.3 Some modifications to the stopping rule BKS

The stopping rule BKS may be slightly modified by dropping the restriction that equal number of observations n^* are taken from each population or (and) using the posterior estimates of p_1 and p_2. The new stopping rules are described as follows.

(i) The stopping rule BKS_1

A modification to the stopping rule BKS given in 7.2.1 using the posterior estimates of p_1 and p_2. The stopping rule becomes:
Stop sampling and then proceeds to BKT at the first stage m at which at least one of the two populations π_1 and π_2 satisfy

$$c_i \geq c_j + (n^* - d_j) \times \frac{r_j}{n_j} \quad \text{for all} \quad i \neq j \ (i, j = 1, 2), \tag{7.2.5}$$

where the second term on the right hand side is the current posterior expected number of successes if the remaining $(n^* - n_j)$ trials taken on population j.

(ii) <u>The stopping rule BKS_2</u>

In many practical problems it is not possible to have all the samples of the same size, since in some cases the investigator has no control over the sample sizes. Hence another modification can be made to the stopping rule BKS so that the maximum number of observations on each population is not restricted.

The stopping rule becomes:

Stop sampling and then proceeds to BKT at the first stage m at which at least one of the two populations π_1 and π_2 satisfy

$$c_i \geq c_j + (N - d_1 - d_2) \quad \text{for all} \quad i \neq j \ (i, j = 1, 2), \tag{7.2.6}$$

where the second term on the right hand side represents the number of potential successes if the remaining $(N - d_1 - d_2)$ observations are taken on π_j.

(iii) <u>The stopping rule BKS_3</u>

A modification of the stopping rule BKS_2 using the posterior estimates of p_1 and p_2. The new stopping rule becomes:

Stop sampling and then proceed to BKT at the first stage m at which at least one of the two populations π_1 and π_2 satisfy

$$c_i \geq c_j + (N - d_1 - d_2) \times \frac{r_j}{n_j} \quad \text{for all} \quad i \neq j \ (i, j = 1, 2), \tag{7.2.7}$$

where the second term on the right hand side represents the current posterior expected number of successes if the remaining

$(N - d_1 - d_2)$ observations are taken on π_j.

7.2.4 Performance characteristics of BKR and BKR* with BKS and modifications

The values of the measures were calculated from the results of MC simulations with 10000 trials in two cases, for generated values of (p_1, p_2) from a prior distribution (uniform in the case we considered) and for fixed values of (p_1, p_2). Table (7.26) shows the accuracy of our simulation results compared with exact results given in Bechhofer and Kulkarni (1982). It should be noted that the measures are related to N since this appears in the stopping rules.

In the following we discuss in details the performance of the sampling rules BKR and BKR* in conjunction with stopping rules BKS, BKS_1, BKS_2 and BKS_3 and the terminal decision rule BKT.

The MC estimates of P(CS)

Table (7.12) and (7.13) give the performance characteristics of the designs that consists of the sampling rules BKR and BKR* in conjunction with the stopping rules BKS, BKS_1, BKS_2 and BKS_3 for generated p_1 and p_2. For all designs, P(CS) increases as N increases and varies little except where the conjugate sampling rule BKR* is used in conjunction with the stopping rules with unrestricted maximum sample sizes that is BKS_2 and BKS_3 in which (see Table 7.13) the P(CS) decreases as N increases. In general, the larger P(CS) are associated with larger E(M); however the designs (BKR, BKS_1) and (BKR, BKS_3) have noticeably smaller E(M).

For all designs; the index P(CS)/E(M) is a decreasing function of N. However, the sampling rules BKR and BKR* give identical results under the stopping rule BKS while the values of this ratio are higher for BKR than those for BKR* under other stopping rules.

Tables (7.14 - 7.25) demonstrate the performance characteristics where p_1 and p_2 are fixed. Table (7.14) and (7.15) show that BK and BK* achieve nearly the same P(CS) uniformly in (p_1, p_2) and hence this numerically supports the property (1) in subsection 7.2.2. Further, it is apparent from these tables that

$$P(CS|p_{[1]} = h,\ p_{[2]} = h + \Delta') = P(CS|p_{[1]} = 1 - h - \Delta',\ p_{[2]} = 1 - h),\quad h, \Delta' > 0,\ h + \Delta' < 1,$$

where BKR and BKR* are used with BKS. This symmetric behaviour is not true under other stopping rules. As in the case where p_1 and p_2 are generated, P(CS) increases as N increases except the scheme (BKR*, BKS_2) in which P(CS) decreases uniformly in (p_1, p_2) while (BKR*, BKS_3) decreases for $p_{[1]} \geq .3$ as N increases. It is noted that P(CS) for (BKR*, BKS_1) is higher than that for (BKR, BKS_1) where $N \geq 20$ uniformly in (p_1, p_2).

In terms of the ratio P(CS)/E(M), the values of (BKR, BKS) where $p_{[1]} = h$, $p_{[2]} = h + \Delta'$ is equal to the values of (BKR*, BKS) where $p_{[1]} = 1 - h - \Delta'$, $p_{[2]} = 1 - h$ and generally BKR is better than BKR* under other stopping rules

For all designs; the index P(CS)/E(M) is a decreasing function of N. However, the sampling rules BKR and BKR* give identical results under the stopping rule BKS while the values of this ratio are higher for BKR than those for BKR* under other stopping rules.

Tables (7.14 - 7.25) demonstrate the performance characteristics where p_1 and p_2 are fixed. Table (7.14) and (7.15) show that BK and BK* achieve nearly the same P(CS) uniformly in (p_1, p_2) and hence this numerically supports the property (1) in subsection 7.2.2. Further, it is apparent from these tables that

$$P(CS|p_{[1]} = h,\ p_{[2]} = h + \Delta') = P(CS|p_{[1]} = 1 - h - \Delta',\ p_{[2]} = 1 - h),$$
$$h,\ \Delta' > 0,\ h + \Delta' < 1,$$

where BKR and BKR* are used with BKS. This symmetric behaviour is not true under other stopping rules. As in the case where p_1 and p_2 are generated, P(CS) increases as N increases except the scheme (BKR*, BKS_2) in which P(CS) decreases uniformly in (p_1, p_2) while (BKR*, BKS_3) decreases for $p_{[1]} \geqslant .3$ as N increases. It is noted that P(CS) for (BKR*, BKS_1) is higher than that for (BKR, BKS_1) where $N \geqslant 20$ uniformly in (p_1, p_2).

In terms of the ratio P(CS)/E(M), the values of (BKR, BKS) where $p_{[1]} = h$, $p_{[2]} = h + \Delta'$ is equal to the values of (BKR*, BKS) where $p_{[1]} = 1 - h - \Delta'$, $p_{[2]} = 1 - h$ and generally BKR is better than BKR* under other stopping rules except for small sample sizes where BKR* is better than BKR under BKS_1.

The MC estimates of E(M)

From Tables (7.12) and (7.13), where p_1 and p_2 are generated, the E(M) for BKR is very close to BKR* under the stopping rule BKS and uniformly better than BKR* under other stopping rules. Substantial reduction in E(M) can be gained using the stopping rules BKS_1 and BKS_3, particularly under the sampling rule BKR. In the scheme (BKR*, BKS_2), the expected sample size to termination will be N - 1, the maximum possible under the stopping rule.

The percent saving [N - E(M)]100/N in the expected sample size is a decreasing function of N and has roughly the same values for both sampling rules BKR and BKR* under the stopping rule BKS while the former sampling rule is uniformly better than other under other stopping rules. More importantly the results show a fairly substantial decrease in E(M) about 66% under the stopping rules which use prior information while a decrease of about 30% is achieved under other stopping rules.

Tables (7.16) and (7.17) contain some results of E(M) where p_1 and p_2 are fixed. It is obvious that as $p_{[1]}$ increases, E(M) decreases under the stopping rules BKS and BKS_2 while the reduction in E(M) takes place with extreme values of p_1 and p_2 under the stopping rules BKS_1 and BKS_3. Further, Table (7.16) shows that the schemes (BKR, BKS) and (BKR, BKS_2) are very close; however, the schemes (BKR, BKS_1) and (BKR, BKS_3), where prior information is used, have noticeably smaller E(M). This suggests that if E(M) is the most important criteria in choosing the scheme then it is worth considering those which use prior information. Moreover, the results in Table (7.16) show that the sampling rule BKR minimizes E(M) for $p_1 + p_2 \geq 1$ and those in Table (7.17) show that the sampling rule BKR* minimizes E(M) for $p_1 + p_2 \leq 1$ under the stopping rule BKS. This provides numerical evidence in support of property (2) given in subsection 7.2.2.

<u>The MC estimates of $E(N_{(1)})$ and $E(N^*_{(1)})$</u>

As can be seen from Table (7.12) and (7.13), where p_1 and p_2 are generated from uniform distribution, the sampling rule BKR achieves substantial reduction in $E(N_{(1)})$ under the stopping rules BKS_1 and BKS_3 whilst the sampling rule BKR* yields high values of $E(N_{(1)})$ under all stopping rules. However, an improvement in the performance measure $E(N_{(1)})$ can be achieved if (BKR*, BKS_1) is used rather than (BKR*, BKS).

The indices $E(N_{(1)})/E(M)$ and $E(N^*_{(1)})/N$ are decreasing (increasing) functions of N under the sampling rule BKR (BKR*) for all stopping rules. The expected number of observations on the poorer population was found to be around (33-35)% ((63-67)%) of the expected sample size, E(M), under the sampling rule BKR (BKR*). Further, the ratio $E(N^*_{(1)})/N$ was found to be around (20-30)% ((40-60)%) under BKR (BKR*).

The results given in Tables (7.18 - 7.21) where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$, show that under the sampling rule BKR, substantial reduction can be gained where prior information is used and it is superior to the sampling rule BKR* under all stopping rules.

The MC estimates of E(R) and E(R*)

In Table (7.13), where p_1 and p_2 are generated, if E(M) is taken into account then the sampling rule BKR* yields low values of E(R) and E(R*) compared with the values produced by the sampling rule BKR, given in Table (7.12), under all stopping rules.

It would be reasonable to compare the schemes in terms of the indices E(R)/E(M) and E(R*)/N which indicate the superiority of the sampling rule BKR over the sampling rule BKR* under all stopping rules.

Some results for E(R) and E(R*) with fixed p_1 and p_2 are presented in Tables (7.22 - 7.25). It is obvious that these performance measures increase as $p_{[1]}$ increases ($p_{[2]} - p_{[1]} = .2$). Furthermore, the values of E(R)/E(M) based on the results given in Tables (7.22 - 7.23) and the results of E(M) given in Tables (7.16 - 7.17) also show the superiority of the sampling rule BKR over the sampling rule BKR* under all stopping rules. In addition, the values of the index E(R*)/N, based on the results given in Tables (7.24 - 7.25), demonstrate the same trend as that shown by the index E(R)/E(M).

The conclusion to be drawn from the results given in this section is that the schemes (BKR, BKS) and (BKR, BKS_2) should be used if P(CS) is the more important criterion while the schemes (BKR, BKS_1) and (BKR, BKS_3) are preferred if E(M) or $E(N_{(1)})$ are of more interest.

The difference between the stopping rule BKS and BKS_2 are minimal when used with the stopping rule BKR; in addition, restricting the maximum sample size on each population to N/2 seems to have little effect in these cases.

Table (7.12)

Performance characteristics of the schemes consisting of the sampling rule BKR in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are generated from a uniform distribution.

Stopping rule	N	Performance characteristics					
		P(CS)	E(M)	$E(N_{(1)})$	E(R)	$E(N^*_{(1)})$	$E(R^*)$
BKS	10	.8110	6.8588	2.5056	3.7014	2.9510	5.4405
	20	.8686	14.3207	4.9012	7.8322	5.3526	11.5775
	30	.8877	21.7654	7.3446	12.0069	7.8383	17.7308
	40	.9053	29.1782	9.7516	16.1738	10.2502	23.9372
	50	.9076	36.7129	12.3156	20.3592	12.8684	30.0408
BKS_1	10	.7098	8.2679	1.2506	1.8322	3.2040	5.3546
	20	.7974	6.5348	2.2829	3.6997	4.8929	11.8049
	30	.8502	10.2141	3.4169	5.8108	6.1165	18.3699
	40	.8680	13.4558	4.4121	7.7562	7.5639	24.7708
	50	.8825	16.9056	5.6196	9.7267	8.9099	31.2499
BKS_2	10	.8025	7.0706	2.5161	3.9063	2.9550	5.4858
	20	.8663	14.4815	4.9339	7.9635	5.3787	11.5753
	30	.8803	21.8636	7.3976	12.0596	7.9300	17.7104
	40	.9058	29.2414	9.7171	16.3120	10.1898	24.0393
	50	.9143	36.6524	12.1456	20.4762	12.6391	30.2289

Table (7.12) continued

Performance characteristics of the schemes consisting of the sampling rule BKR in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are generated from a uniform distribution.

Stopping rule	N	Performance characteristics					
		P(CS)	E(M)	$E(N_{(1)})$	E(R)	$E(N^*_{(1)})$	$E(R^*)$
BKS_3	10	.7602	4.6635	1.6811	2.5985	2.9164	5.4439
	20	.8335	9.1790	3.0960	5.1623	4.6544	11.8874
	30	.8650	13.5943	4.4860	7.8019	6.3204	18.2975
	40	.8876	18.2571	6.0130	10.4499	7.8639	24.7738
	50	.8978	23.0070	7.6620	13.2145	9.5872	31.1477

Table (7.13)

Performance characteristics of the schemes consisting of the sampling rule BKR* in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are generated from a uniform distribution.

Stopping rule	N	Performance characteristics					
		P(CS)	E(M)	$E(N_{(1)})$	E(R)	$E(N^*_{(1)})$	$E(R^*)$
BKS	10	.8121	6.8773	4.3515	3.1836	4.7974	4.5446
	20	.8695	14.3519	9.4020	6.5417	9.8727	9.5921
	30	.8932	21.7111	14.4208	9.7005	14.9063	14.4988
	40	.8983	29.1635	19.4096	12.8921	19.9337	19.3899
	50	.9115	36.6782	24.4388	16.2499	24.9467	24.3668
BKS_1	10	.6936	2.8482	1.7210	1.3770	3.8938	5.0360
	20	.8229	8.5755	5.4365	4.3081	7.2848	10.7393
	30	.8662	15.5473	10.0248	7.7190	11.5842	16.0474
	40	.8811	21.6936	14.0238	10.7351	15.6859	21.4549
	50	.9003	29.0595	18.8885	14.2826	20.3165	26.8141
BKS_2	10	.5211	9.0000	6.1097	3.7397	6.5886	3.7397
	20	.4578	19.0000	13.1822	7.8732	13.7244	7.8732
	30	.4293	29.0000	20.2267	11.8596	20.7974	11.8596
	40	.4159	39.0000	27.0868	15.9763	27.6709	15.9763
	50	.4074	49.0000	34.1687	20.0146	34.7613	20.0146

Table (7.13)

Performance characteristics of the schemes consisting of the sampling rule BKR* in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are generated from a uniform distribution.

Stopping rule	N	Performance characteristics					
		P(CS)	E(M)	$E(N_{(1)})$	E(R)	$E(N^*_{(1)})$	$E(R^*)$
BKS	10	.8121	6.8773	4.3515	3.1836	4.7974	4.5446
	20	.8695	14.3519	9.4020	6.5417	9.8727	9.5921
	30	.8932	21.7111	14.4208	9.7005	14.9063	14.4988
	40	.8983	29.1635	19.4096	12.8921	19.9337	19.3899
	50	.9115	36.6782	24.4388	16.2499	24.9467	24.3668
BKS_1	10	.6936	2.8482	1.7210	1.3770	3.8938	5.0360
	20	.8229	8.5755	5.4365	4.3081	7.2848	10.7393
	30	.8662	15.5473	10.0248	7.7190	11.5842	16.0474
	40	.8811	21.6936	14.0238	10.7351	15.6859	21.4549
	50	.9003	29.0595	18.8885	14.2826	20.3165	26.8141
BKS_2	10	.5211	9.0000	6.1097	3.7397	6.5886	3.7397
	20	.4578	19.0000	13.1822	7.8732	13.7244	7.8732
	30	.4293	29.0000	20.2267	11.8596	20.7974	11.8596
	40	.4159	39.0000	27.0868	15.9763	27.6709	15.9763
	50	.4074	49.0000	34.1687	20.0146	34.7613	20.0146

Table (7.13) continued

Performance characteristics of the schemes consisting of the sampling rule BKR* in conjunction with the stopping rules BKS, BKS$_1$, BKS$_2$, BKS$_3$ and the terminal decision rule BKT where p_1 and p_2 are generared from a uniform distribution.

Stopping rule	N	Performance characteristics					
		P(CS)	E(M)	$E(N_{(1)})$	E(R)	$E(N^*_{(1)})$	$E(R^*)$
BKS$_3$	10	.7029	6.0514	3.8682	2.8210	4.8212	4.5554
	20	.6745	14.8188	9.8317	6.8038	10.8063	9.1764
	30	.6656	24.0634	16.0960	10.8681	17.0840	13.5913
	40	.6526	33.5320	22.6137	14.9528	23.6069	17.9119
	50	.6553	43.2553	29.2272	19.0826	30.2091	22.1811

Table (7.14)

P(CS) of the schemes consisting of the sampling rule BKR in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS	10	.7761	.7413	.7367	.7778
	20	.8727	.8241	.8201	.8753
	30	.9166	.8632	.8729	.9118
	40	.9476	.9030	.9063	.9463
	50	.9656	.9249	.9266	.9636
BKS_1	10	.6074	.6216	.6443	.6684
	20	.7482	.7171	.7190	.7526
	30	.8213	.7819	.7869	.8359
	40	.8583	.8215	.8343	.8806
	50	.8891	.8595	.8782	.9165
BKS_2	10	.8741	.7967	.7761	.7957
	20	.9176	.8519	.8341	.8783
	30	.9444	.8823	.8906	.9164
	40	.9619	.9178	.9153	.9492
	50	.9754	.9365	.9337	.9636

Table (7.14) continued

P(CS) of the schemes consisting of the sampling rule BKR in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS_3	10	.7012	.6505	.6699	.7099
	20	.7594	.7531	.7762	.8047
	30	.8410	.8174	.8331	.8731
	40	.8996	.8705	.8834	.9222
	50	.9257	.8989	.9162	.9473

Table (7.15)

P(CS) of the schemes consisting of the sampling rule BKR* in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS	10	.7813	.7358	.7394	.7840
	20	.8702	.8176	.8210	.8694
	30	.9169	.8653	.8632	.9170
	40	.9444	.9035	.9041	.9498
	50	.9655	.9222	.9250	.9632
BKS_1	10	.6056	.6199	.6332	.6328
	20	.7604	.7418	.7605	.8291
	30	.8387	.8162	.8380	.9025
	40	.8747	.8670	.8794	.9369
	50	.9074	.9032	.9084	.9576
BKS_2	10	.8674	.7659	.7276	.6929
	20	.8300	.7114	.6799	.6335
	30	.8023	.6936	.6487	.6131
	40	.7785	.6678	.6309	.5978
	50	.7688	.6570	.6156	.5875

Table (7.15) continued

P(CS) of the schemes consisting of the sampling rule BKR^* in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS_3	10	.7421	.6552	.5707	.4501
	20	.8184	.6678	.4582	.2141
	30	.8376	.6185	.3854	.1244
	40	.8459	.5823	.3374	.0711
	50	.8459	.5551	.3042	.0513

Table (7.16)

E(M) of the schemes consisting of the sampling rule BKR in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS	10	8.3238	7.7206	6.9545	6.0897
	20	17.4397	16.3999	15.0009	12.8554
	30	26.4513	25.1274	23.1136	19.7109
	40	35.4348	33.8210	31.2920	26.6896
	50	44.3752	42.5235	39.4557	33.2123
BKS_1	10	3.2030	3.3729	3.4234	3.2845
	20	7.0861	6.8608	6.9459	6.8135
	30	10.9376	11.5571	11.7346	10.8182
	40	13.9356	15.4678	16.1126	15.0668
	50	17.0857	20.2319	21.2798	19.4006
BKS_2	10	8.4339	7.8714	7.1956	6.3973
	20	17.4887	16.4849	15.1202	13.1610
	30	26.4724	25.1952	23.1886	19.9169
	40	35.4490	33.8618	31.3686	26.6725
	50	44.3755	42.5802	39.4690	33.3697

Table (7.16) continued

E(M) of the schemes consisting of the sampling rule BKR in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS_3	10	5.0128	4.7838	4.7865	4.6305
	20	10.3747	10.2945	10.0163	9.0283
	30	14.6915	15.6811	15.7347	14.2200
	40	19.5409	21.7177	21.8646	19.7286
	50	23.9204	27.7276	28.4768	25.0033

Table (7.17)

E(M) of the schemes consisting of the sampling rule BKR* in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS	10	6.0844	6.9687	7.7147	8.3312
	20	12.9088	15.0253	16.3508	17.4021
	30	19.6474	23.1610	25.1124	26.4445
	40	26.4152	31.3165	33.8298	35.4002
	50	33.2149	39.4664	42.5379	44.3336
BKS_1	10	2.2631	2.6580	3.1008	3.5464
	20	5.7051	7.9016	10.8628	13.8747
	30	10.1432	15.6864	20.8728	23.6440
	40	13.5973	23.0055	29.6058	32.6629
	50	18.8652	32.4332	38.9305	41.7326
BKS_2	10	9.0000	9.0000	9.0000	9.0000
	20	19.0000	19.0000	19.0000	19.0000
	30	29.0000	29.0000	29.0000	29.0000
	40	39.0000	39.0000	39.0000	39.0000
	50	49.0000	49.0000	49.0000	49.0000

Table (7.17) continued

E(M) of the schemes consisting of the sampling rule BKR* in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS_3	10	5.0352	5.7286	6.7864	7.8324
	20	11.3594	15.1680	17.4770	18.3842
	30	18.9862	25.7010	27.8094	28.5028
	40	27.9746	36.1915	37.9339	38.5342
	50	37.5494	46.4674	47.9814	48.5513

Table (7.18)

$E(N_{(1)})$ of the schemes consisting of the sampling rule BKR in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS	10	3.6962	3.3149	2.8231	2.1755
	20	7.6539	6.8680	5.7559	3.8569
	30	11.5990	10.5285	8.6991	5.5494
	40	15.5319	14.1256	11.7516	7.0675
	50	19.4417	17.7683	14.8255	8.6471
BKS_1	10	1.4208	1.4698	1.4438	1.2961
	20	3.0320	2.8346	2.7092	2.3121
	30	4.5917	4.7054	4.4391	3.2821
	40	5.8928	6.2809	5.9343	4.1928
	50	7.2217	8.2397	7.7878	5.2561
BKS_2	10	3.7282	3.3372	2.8519	2.1951
	20	7.6767	6.9020	5.7809	3.8892
	30	11.6035	10.5691	8.6811	5.5211
	40	15.5401	14.1295	11.7686	7.0816
	50	19.4329	17.8005	14.8109	8.6794

Table (7.18) continued

$E(N_{(1)})$ of the schemes consisting of the sampling rule BKR in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS_3	10	2.1433	2.0430	1.9304	1.7199
	20	4.3675	4.2009	3.7532	2.8666
	30	6.2434	6.3702	5.7900	4.0294
	40	8.2394	8.7939	7.9516	5.3032
	50	10.1553	11.2732	10.3810	6.5266

Table (7.19)

$E(N_{(1)})$ of the schemes consisting of the sampling rule BKR^* in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS	10	3.9351	4.1400	4.4018	4.6321
	20	9.0313	9.2638	9.5187	9.7832
	30	14.1916	14.3845	14.6089	14.8507
	40	19.3886	19.5343	19.7021	19.9050
	50	24.5898	24.6037	24.7709	24.9286
BKS_1	10	1.2881	1.4952	1.7430	1.9788
	20	3.6608	4.7994	6.3519	8.0033
	30	6.9236	9.7228	12.3658	13.6333
	40	9.5393	14.4121	17.5412	18.7530
	50	13.4965	20.4128	23.0011	23.8552
BKS_2	10	6.2309	5.5942	5.3048	5.1649
	20	14.0985	12.0313	11.2570	10.8636
	30	21.8423	18.3753	17.1473	16.5099
	40	29.5538	24.7052	22.9854	22.1503
	50	37.1081	30.9962	28.8692	27.8017

Table (7.19) continued

$E(N_{(1)})$ of the schemes consisting of the sampling rule BKR* in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS_3	10	3.1663	3.4341	3.9219	4.4262
	20	7.8734	9.5193	10.3399	10.4278
	30	13.6723	16.2361	16.4034	16.1188
	40	20.6671	22.8916	22.3346	21.7853
	50	28.0984	29.3890	28.2180	27.4259

Table (7.20)

$E(N^*_{(1)})$ of the schemes consisting of the sampling rule BKR in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS	10	3.9710	3.7898	3.5229	3.0054
	20	7.8299	7.2793	6.4045	4.5925
	30	11.7250	10.8986	9.2240	6.1721
	40	15.6173	14.4038	12.1678	7.4806
	50	19.4966	17.9967	15.1694	8.9597
BKS_1	10	4.1128	3.9824	3.7823	3.5378
	20	6.1530	6.4921	6.3618	5.5876
	30	7.7901	8.5122	8.2112	6.3780
	40	9.3868	10.4516	9.7527	7.1557
	50	10.6624	12.1113	10.9744	7.7661
BKS_2	10	3.8913	3.7211	3.4498	2.9511
	20	7.8018	7.2726	6.4091	4.5986
	30	11.6984	10.9146	9.1523	6.1138
	40	15.6105	14.3763	12.1538	7.4833
	50	19.4768	18.0059	15.1429	8.9839

Table (7.20) continued

$E(N^*_{(1)})$ of the schemes consisting of the sampling rule BKR in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS_3	10	3.5503	3.8148	3.6088	3.2926
	20	6.3136	6.3629	5.8477	4.9424
	30	8.2573	8.6609	7.8722	5.9143
	40	9.8714	10.6914	9.6620	6.6437
	50	11.6551	12.9304	11.6920	7.6263

Table (7.21)

$E(N^*_{(1)})$ of the schemes consisting of the sampling rule BKR* in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS	10	4.7457	4.8295	4.8872	4.8965
	20	9.7431	9.8985	9.9406	9.9635
	30	14.7882	14.9364	14.9665	14.9768
	40	19.8348	19.9600	19.9754	19.9847
	50	24.8826	24.9693	24.9860	24.9915
BKS_1	10	4.3762	4.3188	4.2550	4.2484
	20	7.0662	7.8334	8.3817	8.8523
	30	10.0138	12.0826	13.5271	14.0746
	40	12.7340	16.3547	18.3844	19.0442
	50	15.1885	21.6561	23.6227	24.0519
BKS_2	10	6.3635	5.8283	5.5772	5.4720
	20	14.2685	12.3199	11.5771	11.2301
	30	22.0400	18.6817	17.4986	16.8968
	40	29.7753	25.0374	23.3545	22.5525
	50	37.3393	31.3392	29.2536	28.2142

Table (7.21) continued

$E(N^*_{(1)})$ of the schemes consisting of the sampling rule BKR^* in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]}$) .2.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS_3	10	4.4130	4.8265	5.1353	5.4042
	20	9.2833	10.8651	11.5052	11.6247
	30	15.0978	17.6135	17.6499	17.3964
	40	22.0106	24.3389	23.6344	23.1283
	50	29.3819	30.9073	29.5761	28.7888

Table (7.22)

E(R) of the schemes consisting of the sampling rule BKR in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS	10	1.7786	3.2086	4.3147	5.0530
	20	3.6885	6.8299	9.3561	10.8061
	30	5.6360	10.4404	14.4376	16.6395
	40	7.4985	14.0756	19.5790	22.4368
	50	9.3962	17.7049	24.6606	28.1629
BKS_1	10	.6916	1.4196	2.1220	2.7094
	20	1.5275	2.8840	4.3416	5.6869
	30	2.3680	4.8423	7.3557	9.0940
	40	2.9893	6.4823	10.0921	12.7167
	50	3.6487	8.4385	13.3456	16.4110
BKS_2	10	1.7890	3.2742	4.4820	5.3174
	20	3.6997	6.8529	9.4413	11.0715
	30	5.6331	10.4496	14.4978	16.8344
	40	7.5064	14.0957	19.6086	22.5960
	50	9.4089	17.7199	24.6722	28.3038

Table (7.22) continued

E(R) of the schemes consisting of the sampling rule BKR in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS_3	10	1.0920	2.0114	2.9816	3.8295
	20	2.2450	4.3124	6.2825	7.5710
	30	3.1534	6.5546	9.8676	12.0047
	40	4.1942	9.1022	13.7089	16.7035
	50	5.1373	11.6065	17.8662	21.2038

Table (7.23)

E(R) of the schemes consisting of the sampling rule BKR* in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1. .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS	10	1.0497	2.6611	4.5400	6.5667
	20	2.0625	5.6651	9.5340	13.6825
	30	3.0286	8.6908	14.6473	20.8324
	40	4.0124	11.7410	19.7440	27.8933
	50	5.0132	14.7997	24.8498	34.9327
BKS_1	10	.4311	1.0434	1.8310	2.7965
	20	.9894	3.0054	6.3454	10.8645
	30	1.6596	5.8812	12.1436	18.5386
	40	2.1588	8.6009	17.2138	25.6264
	50	2.9458	12.1440	22.6482	32.8153
BKS_2	10	1.4635	3.4098	5.2659	7.0661
	20	2.8465	7.0897	11.0392	14.9026
	30	4.3278	10.8266	16.8687	22.8004
	40	5.7741	14.5651	22.7320	30.6894
	50	7.2741	18.3346	28.5556	38.5508

Table (7.23) continued

E(R) of the schemes consisting of the sampling rule BKR* in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS_3	10	.8872	2.2038	3.9891	6.1712
	20	1.8264	5.6935	10.1643	14.4362
	30	2.9386	9.6098	16.1887	22.4483
	40	4.2535	13.5081	22.1044	30.3271
	50	5.6322	17.3586	27.9913	38.2206

Table (7.24)

$E(R^*)$ of the schemes consisting of the sampling rule BKR in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS	10	1.8149	3.9192	5.8676	7.8249
	20	3.9392	8.2455	12.2740	16.5669
	30	6.1890	12.5131	18.7198	25.1870
	40	8.3711	16.8299	25.1527	34.0049
	50	10.6152	21.1236	31.5267	42.6346
BKS_1	10	1.7909	3.8988	5.5945	7.8027
	20	4.1836	8.3207	12.3046	16.3134
	30	7.0630	12.9951	19.0229	25.1361
	40	9.7012	17.5893	25.6069	34.0548
	50	12.4303	22.2870	32.4445	42.8693
BKS_2	10	1.8131	3.9152	5.8585	7.7945
	20	3.9452	8.2211	12.2631	16.5910
	30	6.1831	12.4850	18.7312	25.1883
	40	8.3754	16.8285	25.1310	34.0138
	50	10.6304	21.1102	31.5281	42.6298

Table (7.24) continued

$E(R^*)$ of the schemes consisting of the sampling rule BKR in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS_3	10	1.7798	3.9518	6.0685	7.9490
	20	4.3557	8.4120	12.5009	16.4670
	30	6.9616	13.0324	18.9867	25.3243
	40	9.5486	17.5850	25.5718	34.2185
	50	12.2249	22.1508	32.2270	42.9170

Table (7.25)

$E(R^*)$ of the schemes consisting of the sampling rule BKR^* in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS	10	1.5967	3.6970	5.6187	7.2347
	20	3.6570	7.7276	11.5517	15.2792
	30	5.5521	11.7135	17.5321	23.3817
	40	7.5605	15.7242	23.5558	31.4678
	50	9.5354	19.7271	29.5822	39.5098
BKS_1	10	1.6225	3.8852	5.7190	7.7989
	20	4.0743	8.1353	11.9573	15.7107
	30	6.4923	12.2933	17.8586	23.7323
	40	9.0045	16.4281	23.8840	31.7324
	50	11.3183	20.4089	29.8316	39.7213
BKS_2	10	1.4635	3.4098	5.2659	5.4720
	20	2.8465	7.0897	11.0392	11.2301
	30	4.3278	10.8266	16.8687	16.8968
	40	5.7741	14.5651	22.7320	22.5525
	50	7.2741	18.3346	28.5556	28.2142

Table (7.25) continued

$E(R^*)$ of the schemes consisting of the sampling rule BKR^* in conjunction with the stopping rules BKS, BKS_1, BKS_2, BKS_3 and the terminal decision rule BKT where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

Stopping rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKS_3	10	1.5493	3.6631	5.6938	7.3199
	20	3.6814	7.4702	11.3952	15.0499
	30	5.5073	11.0703	17.1725	22.9455
	40	7.0808	14.6908	22.9944	30.7929
	50	8.6296	18.3847	28.8360	38.6693

Table (7.26)

Comparison of approximate and exact+ results of the performance characteristics P(CS), E(M) and $E(N_{(1)})$ for the scheme BK where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

$(p_{[1]}p_{[2]})$	N	P(CS) Approx.	P(CS) Exact	E(N) Approx.	E(N) Exact	$E(N_{(1)})$ Approx.	$E(N_{(1)})$ Exact
(1., .3)	10	.7761	.7786	8.3238	8.345	3.6962	3.710
	20	.8727	.8697	17.4397	17.430	7.6539	7.650
	30	.9166	.9181	26.4513	26.433	11.5990	11.578
	40	.9476	.9470	35.4348	35.397	15.5319	15.494
	50	.9656	.9650	44.3752	44.336	19.4417	19.402
(.3, .5)	10	.7413	.7374	7.7206	7.709	3.3149	3.307
	20	.8241	.8188	16.3999	16.406	6.8680	6.881
	30	.8632	.8688	25.1274	25.114	10.5285	10.492
	40	.9030	.9027	33.8210	33.805	14.1256	14.104
	50	.9249	.9268	42.5235	42.474	17.7683	17.711
(.5, .7)	10	.7367	.7374	6.9545	6.970	2.8231	2.821
	20	.8201	.8188	15.0009	15.018	5.7559	5.761
	30	.8729	.8688	23.1136	23.169	8.6991	8.766
	40	.9063	.9027	31.2920	31.321	11.7516	11.797
	50	.9266	.9268	39.4557	39.454	14.8255	14.831

Table (7.26) continued

Comparison of approximate and exact[+] results of the performance characteristics P(CS), E(M) and $E(N_{(1)})$ for the scheme BK where p_1 and p_2 are fixed with $p_{[2]} - p_{[1]} = .2$.

$(p_{[1]} p_{[2]})$	N	P(CS)		E(N)		$E(N_{(1)})$	
		Approx.	Exact	Approx.	Exact	Approx.	Exact
(.7, .9)	10	.7778	.7786	6.0897	6.088	2.1755	2.168
	20	.8753	.8697	12.8554	12.868	3.8569	3.878
	30	.9118	.9181	19.7109	19.670	5.5494	5.457
	40	.9463	.9470	26.4896	26.438	7.0675	7.015
	50	.9636	.9650	33.2123	33.175	8.6471	8.587

\+ The axact results are abstracted from Tables (4.4A, 4.4B, 4.6) in Bechhofer and Kulkarni (1982).

7.3 Some further selection schemes

In this section we discuss some further schemes which are combinations of the various sampling, stopping and terminal decision rules already given in the previous sections. Let GS denotes the grop sequential sampling rule where two observations, one on each population, are taken at a time. The investigation also covers three further sampling rules, these are:

(i) Play-The-Winner sampling rule (PWR)

This sampling rule was first suggested by Robbins (1956) in a discussion of the two armed-bandit problem (Büringer et al. (1980)) and its application to the problem of allocating observations among treatments appeared in Zelen (1969). According to this rule the observations are taken sequentially one at a time. At the outset one of the populations is randomly selected and the first observation is taken from this population. If there exists some prior information on the populations then the first observation is taken from the population which has the largest prior mean, or in the event of a tie, at random. In subsequent stages a success generates another trial on the same population and a failure causes a switch to the other population. It is assumed that each observation has instantaneous response. The use of PWR causes a bias in favour of sampling the better population.

(ii) Play-The-Loser sampling rule (PLR)

It is the conjugate of PWR, observations are taken sequentially one at a time and the first observation is taken from one of the population which is chosen randomly at the outset. Contrary to PWR, the first observation is taken from the worse population if some prior information available and in subsequent stages a failure with a given population generates a further trial on the same population whereas a success causes a switch to the other population. The use of PLR causes a bias in favour of sampling from the worse population.

(iii) <u>Play-The-Clear-Winner (PCWR)</u>

This rule was proposed by Arkles and Srinivasan (1979) and is a modification to PWR whereby observations are taken sequentially either from both or one of the populations at each stage. At the first stage one observation is made from each of the populations. In subsequent stages sampling takes place on one or both populations depending on the outcome of the preceding stage. At the i^{th} stage $(i = 2, 3, \dots)$ observations are made on both populations π_1 and π_2 if at the $(i - 1)^{th}$ stage, either

(a) π_1 and π_2 were both observed and the results were both successes or both failures, or

(b) either π_1 or π_2 was observed and the result was a failure.

At the i^{th} stage $(i = 2, 3, \dots)$, π_1 (π_2) is observed if at the $(i - 1)^{th}$ stage either

(a) π_1 and π_2 were observed, π_1 (π_2) resulting in a success and π_2 (π_1) in a failure, or

(b) either π_1 or π_2 was observed and the result was a success.

This formulation implies that PCWR is equivalent to implementing two PWR in parallel, one starting with π_1 and the other with π_2.

In the following we discuss the other combinations of sampling and stopping rules which were not discussed in the last sections.

(1) The sampling rules BKR, BKR*, PWR, PLR, PCWR under the stopping rule DS

In Tables (7.27), where p_1 and p_2 are generated, the sampling rules BKR, BKR*, PWR, PLR and PCWR achieve roughly the same P(CS) if used with the stopping rule DS and the terminal decision rule DT. The same table also shows that PCWR has slightly smaller E(M) than others; in addition, based on the indices P(CS)/E(M) and E(M)/N, the performance of the sampling rules can be ordered as

$$\text{PCWR} > [\text{BKR} \underset{-}{\sim} \text{BKR}^* \underset{-}{\sim} \text{PWR} \underset{-}{\sim} \text{PLR}],$$

where large (small) values of the first (second) index is preferable.

It is intuitively clear and strongly supported by numerical results of Table (7.27) that the sampling rules which have the property of stay-on-the-winner (BKR, PWR, PCWR) have smaller $E(N_{(1)})$ than the sampling rules BKR* and PLR which have the property of stay-on-the-loser. The ratio $E(N_{(1)})/E(M)$ is an increasing function of N under the sampling rules BKR, PWR and is a decreasing function of N under other sampling rules given in Table (7.27). The measure $E(N^*_{(1)})$ has the same behaviour as $E(N_{(1)})$; however, the ratio $E(N^*_{(1)})/N$ is a decreasing function of N for all schemes and the performance of the sampling rules under this ratio can be ordered as follows:

$$BKR > PWR > PCWR > PLR > BKR^*.$$

It is obvious from Table (7.27) that sampling rules BKR and PWR have larger values of E(R) than those for other sampling rules for small N ($N \leq 30$). Furthermore, BKR and PWR are better than others in terms of $E(R^*)$ for all N. The ratio E(R)/E(M) is a decreasing function of N under the sampling rules BKR, PWR, PCWR whilst it is an increasing function of N under BKR^* and PLR. Moreover, the ratio $E(R^*)/N$ is an increasing function of N under all sampling rules and the performance of the sampling rules measured by this ratio can be ordered as

$$[BKR \simeq PWR \simeq PCWR] > [BKR^* \simeq PLR].$$

Tables (7.28 - 7.33) present some results for fixed p_1 and p_2 with $p_{[2]} - p_{[1]} = .2$. Table (7.28) shows that P(CS) is higher for extreme values of $p_{[1]}$ than those for moderate under all sampling rules. It seems also from this table that BKR, PWR, PCWR (BKR^*, PLR) are preferable if $p_{[1]}$ is small (large).

The results of E(M) given in Table (7.29) show that the sampling rule PCWR is the best for small and moderate values of p_1, p_2 and BKR is the best for very large values of p_1, p_2. However, BKR^* and PLR have similar values of E(M). In addition, the sampling rules BKR, PWR and PCWR are superior (inferior) to the sampling rules BKR^* and PLR for large (small) values of $p_{[1]}$.

In terms of P(CS)/E(M), the sampling rules BKR^* and PLR perform better than the sampling rules BKR, PWR and PCWR for small values of $p_{[1]}$ ($p_{[1]} \leq .3$).

All the results in Table (7.30) indicate the superiority of BKR, PWR and PCWR over BKR* and PLR if the performance is measured by $E(N_{(1)})$, provided that $p_{[1]}$ is large. The superiority of BKR, PWR and PCWR is obvious under the measure $E(N^*_{(1)})$ (see Table (7.31)).

The results in Table (7.32) show that E(R) is an increasing function of $p_{[1]}$ and indicate that neither sampling rule leads to uniformly better performance for all $p_{[1]}$. Rather there is a crossover point for $p_{[1]}$ above which BKR* and PLR are the better performers; below this point BKR, PWR and PCWR are the better performers.

From Table (7.33) we note that the rules BKR, PWR and PCWR are uniformly better than BKR* and PLR as far as $E(R^*)$ is concerned.

The ratio E(R)/E(M) has roughly constant values under each pair of values of $p_{[1]}$, $p_{[2]}$ for each sampling rule and for all N. The values of this ratio also indicate the superiority of BKR, PWR and PCWR over BKR* and PLR.

The ratio $E(R^*)/N$ is an increasing function of N and shows also the uniform superiority of BKR, PWR and PCWR over BKR* and PLR.

(2) <u>The sampling rules A_2, A_5, GS, PWR, PLR, PCWR under the stopping rule BKS</u>

Table (7.34) presents some numerical results on the performance characteristics where p_1 and p_2 are generated from a uniform distribution.

It follows from a theorem in Jennison (1983) that these sampling rules achieve equal P(CS) if used with the stopping rule BKS and the terminal decision rule BKT and the results of Table (7.34) confirm this property; the small variations are due to simulation fluctuations.

All the sampling rules have generally similar E(M), except GS which has slightly larger values of E(M). The percent reduction in E(M) is a fairly constant (about 23%) in GS for all N; while it is a decreasing function of N and within the range (25 - 31)% as N goes from 10 to 50 for other sampling rules.

Under the measures $E(N_{(1)})$ and $E(N^*_{(1)})$ we found that the performance is as follows:

$$[A_5 \simeq PWR \simeq PCWR] > GS > [A_2 \simeq PLR].$$

The ratio $E(N_{(1)})/E(M)$ is greater than ½ and has similar values under A_2, PLR and less than ½ and has similar values under BKR, PWR, A_5.

The measures E(R) and $E(R^*)$ have roughly the same performance ordering as that for $E(N_{(1)})$ and $E(N^*_{(1)})$.

Tables (7.35 - 7.40) show the performance characteristics for fixed p_1 and p_2 with $p_{[2]} - p_{[1]} = .2$.

Again, the results of P(CS) given in Table (7.35) show that these sampling rules have the same P(CS) uniformly in (p_1, p_2).

From Table (7.36) we note that E(M) is an increasing (a decreasing) function of $p_{[1]}$ under the rules A_2, PLR (A_5, PWR, PCWR) and roughly constant under GS. A_2 and A_5 are very similar to PLR and PWR, PCWR respectively.

It is apparent from Tables (7.37) and (7.38) that the performance of the sampling rules as measured by $E(N_{(1)})$ and $E(N^*_{(1)})$ roughly has the same ordering as that given under the case where p_1 and p_2 are generated. Each of the sampling rules A_2 and PWR has roughly constant $E(N_{(1)})$ and $E(N^*_{(1)})$ for all values of $p_{[1]}$, $p_{[2]}$ with $p_{[2]} - p_{[1]} = .2$.

Under the measures $E(R)$ and $E(R^*)$, the results of which are given in Tables (7.39) and (7.40) respectively, the performance of the sampling rules follow the same ordering as that given for $E(N_{(1)})$ and $E(N^*_{(1)})$ above.

(3) <u>The sampling rules A_2, A_5, GS, PWR, PLR, PCWR under the stopping rule BKS_1</u>

Table (7.41) gives the performance measures of these schemes for generated p_1 and p_2.

Under the stopping rule BKS_1 and the terminal decision rule BKT, the P(CS) is not the same for these sampling rules; therefore the comparison of these sampling rules using the index P(CS)/E(M) will be more sensible. The values of this ratio show that the performance of the sampling rules can be ordered as

$$[PLR \simeq A_2] > [A_5 \simeq PWR] > PCWR > GS \qquad \text{for small } N \ (N \leqslant 10),$$

and

$$[A_5 \simeq PWR] > PCWR > GS > [A_2 \simeq PLR] \qquad \text{for large } N \ (N > 10).$$

The results of $E(N_{(1)})$ indicate that the performance of the sampling rules can be ordered as follows:

$$A_5 > PWR > PCWR > GS > PLR > A_2.$$

It is noted that the ratio $E(N_{(1)})/E(M)$ is a decreasing (an increasing) function of N and less than (greater than) ½ under the sampling rules A_5, PWR, PCWR (A_2, PLR).

Furthermore, the ratio $E(R)/E(M)$ is an increasing (a decreasing) function of N and greater than (less than) ½ for the sampling rules A_5, PWR, PCWR (A_2, PLR). However, $E(N^*_{(1)})$ and $E(R^*)$ provide more direct comparison for the sampling rules since they are functions of P(CS), E(M) and $E(N_{(1)})$. According to these two measures, the performance of the sampling rules has the following ordering:

$$[A_5 \simeq PWR \simeq PCWR] > GS > PLR > A_2.$$

Results of the performance characteristics of the sampling rules for fixed p_1 and p_2 with $p_{[2]} - p_{[1]} = .2$ are given in Tables (7.42 - 7.47).

As it is clear from Table (7.42), the P(CS) are different for different sampling rules; therefore, $E(N^*_{(1)})$ and $E(R^*)$ are reasonable measures to be used. Under $E(N^*_{(1)})$ the performance ordering roughly is

$$[A_5 \simeq PWR \simeq PCWR] > GS > PLR > A_2,$$

and in terms of $E(R^*)$ the ordering becomes

$$[A_5 \simeq PWR \simeq PCWR] > GS > A_2 > PLR.$$

(4) The sampling rules A_2, A_5, GS, PWR, PLR, PCWR under the stopping rule BKS_2

Some numerical results concerning the performance of the sampling rules A_2, A_5, GS, PWR, PLR, PCWR in conjunction with the stopping rule BKS_2 and the terminal decision rule BKT for generated p_1 and p_2 are presented in Table (7.48).

It can be observed from this table that P(CS) is an increasing function of N under all sampling rules excpet A_2. Using P(CS), the ordering of the performance is

$$[GS \simeq PWR \simeq PCWR] > A_5 > PLR > A_2$$

and using E(M) is

$$A_5 > PWR > PCWR > A_2 > GS > PLR.$$

From the above statement it is not obvious which sampling rule is better since different P(CS) corresponding different E(M); therefore, the ratio P(CS)/E(M) might be chosen as a reasonable alternative in this situation. We found that this ratio is a decreasing function of N under all sampling rules and the performance as measured by this ratio can be ordered as

$$A_5 > PWR > PCWR > GS > PLR > A_2 ,$$

with A_5 is slightly superior and A_2 is drastically inferior.

More direct and appropriate comparison can be made using $E(N^*_{(1)})$ and $E(R^*)$. Under these measures, we found the performance is the same as that under P(CS)/E(M).

The performance of these schemes are also investigated for fixed p_1 and p_2 with $p_{[2]} = p_{[1]} + .2$; the results are given in Tables (7.49 - 7.54).

The results given in Tables (7.49) and (7.50) show clearly that the ratio P(CS)/E(M) is a decreasing function of N for all sampling rules; further, it is found that PWR is superior and A_2 is inferior to others if this ratio is considered.

Based on the results of $E(R^*)$ given in Table (7.54) we observe that the performance ordering is as follows:

$$A_5 > [\text{PWR} \simeq \text{PCWR}] > \text{GS} > \text{PWR} > A_2.$$

(5) <u>The sampling rules A_2, A_5, GS, PWR, PLR, PCWR under the stopping rule BKS_3</u>

Table (7.55) contains some results on the performance characteristics of the sampling rules A_2, A_5, GS, PWR, PLR, PCWR in conjunction with the stopping rule BKS_3 and the terminal decision rule BKT. Under BKS_3 again we have different P(CS) and different E(M); hence the index P(CS)/E(M) is more useful to evaluate the schemes. According to this ratio we can order the performance of the sampling rules as

$$A_5 > \text{PWR} > \text{PCWR} > \text{GS} > \text{PLR} > A_2.$$

It is also clear that

$$A_5 > \text{PWR} > \text{PCWR} > \text{GS} > [A_2 \simeq \text{PLR}]$$

if the performance is measured by $E(N_{(1)})$.

The ratio $E(N_{(1)})/E(M)$ is an increasing (a decreasing) function of N under A_2, PLR (A_5, PCWR, PWR) and A_5 was found to be superior to others using this ratio.

Under the measures $E(R)/E(M)$, $E(N_{(1)})$ and $E(N^*_{(1)})$, the performance of the sampling rules has the same ordering as that given for $P(CS)/E(M)$. In terms of $E(R^*)$, the performance ordering is as follows:

$$A_5 > PWR > PCWR > GS > PLR > A_2.$$

Results for the cases where p_1 and p_2 are fixed with $p_{[2]} = p_{[1]} + .2$ are given in Tables (7.56 - 7.61). Again, here we have different P(CS) for different sampling rules; therefore a measure such as $E(R^*)$ will be more helpful. We found that the performance of the sampling rules based on this measure can be ordered as follows

$$[A_5 \simeq PWR \simeq PCWR] > GS > PLR > A_2.$$

From this set of results we draw the following conclusions.

(1) Under the stopping rule DS, we note that the sampling rule PCWR, PWR and BKR are very close in their performance and superior to others. If we take into account the ease of use then PCWR might be preferred.

(2) Under the stopping rule BKS and BKS_1, the sampling rules A_5, PWR and PCWR are superior to others and no significant differences between them are detected.

(3) Under the stopping rule BKS_2 and BKS_3, the sampling rule A_5 is marginally superior to others.

Table (7.27)

Performance characteristics of the sampling rules BKR, BKR*, PWR, PLR and PCWR under the stopping rule DS and the terminal decision rule DT for generated p_1, p_2, where N = 10, 30, 50, δ_0 = .4 with uniform priors.

Sampling rule	N	Performance characteristics					
		P(CS)	E(M)	$E(N_{(1)})$	E(R)	$E(N^*_{(1)})$	$E(R^*)$
BKR	10	.8154	8.3688	2.9362	4.6808	3.0854	5.6440
	30	.8767	17.8937	6.4714	9.1992	7.4266	18.2090
	50	.8952	15.9359	9.6560	12.9429	11.4536	30.9816
BKR*	10	.8167	8.2119	5.2799	3.6887	5.4563	4.6086
	30	.8775	17.4674	11.1196	8.5070	12.1382	16.4090
	50	.8965	26.0057	16.2843	13.3424	18.1878	28.6737
PWR	10	.8223	8.3480	3.0046	4.6449	3.1554	5.6105
	30	.8751	17.8512	6.6845	9.2069	7.5405	18.1507
	50	.8988	25.9759	9.8879	12.9403	11.6853	30.8106
PLR	10	.8161	8.2155	5.2250	3.6792	5.4038	4.6002
	30	.8702	17.6593	10.9969	8.7531	12.0764	16.4396
	50	.8922	25.5191	15.7446	13.0185	17.6648	28.6818
PCWR	10	.8047	7.4163	3.1162	3.8823	3.3669	5.5640
	30	.8710	16.5982	6.6213	8.5192	7.8147	18.0229
	50	.9020	24.9977	9.8128	12.7029	11.7029	30.8196

Table (7.28)

P(CS) of the sampling rules BKR, BKR*, PWR, PLR and PCWR under the stopping rule DS and the terminal decision rule DT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50, $\delta_0 = .4$ with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKR	10	.8715	.7893	.7810	.7860
	30	.9404	.8725	.8490	.8833
	50	.9700	.9073	.8810	.9041
BKR*	10	.7990	.7735	.7958	.8656
	30	.8805	.8420	.8689	.9415
	50	.9033	.8796	.9155	.9723
PWR	10	.8846	.8038	.7900	.8004
	30	.9486	.8773	.8454	.8903
	50	.9690	.9098	.8898	.9127
PLR	10	.8068	.7841	.8043	.8815
	30	.8810	.8527	.8726	.9422
	50	.9083	.8724	.9171	.9741
PCWR	10	.8879	.8030	.7921	.8504
	30	.9424	.8603	.8352	.8882
	50	.9683	.8969	.8561	.9259

Table (7.29)

E(M) of the sampling rules BKR, BKR*, PWR, PLR and PCWR under the stopping rule DS and the terminal decision rule DT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50, $\delta_0 = .4$ with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKR	10	9.6772	9.0152	8.4415	8.4840
	30	27.6800	23.4392	19.5896	17.4809
	50	44.9824	36.3529	30.0233	25.0859
BKR*	10	8.2058	8.3611	8.9672	9.6387
	30	16.5032	19.5024	23.1111	27.5863
	50	24.0586	29.3430	36.3375	45.3345
PWR	10	9.6093	8.9364	8.4046	8.5307
	30	27.4197	22.7991	19.3411	17.6699
	50	44.7352	36.0001	29.6450	26.0197
PLR	10	8.2214	8.3812	8.9547	9.6278
	30	16.6358	19.4326	22.9573	27.4180
	50	24.2759	29.3444	36.0482	44.7665
PCWR	10	9.2790	8.1141	7.4396	7.8801
	30	26.1658	20.9191	17.8581	19.1807
	50	42.2582	32.4747	27.1861	29.2762

Table (7.30)

$E(N_{(1)})$ of the sampling rules BKR, BKR*, PWR, PLR and PCWR under the stopping rule DS and the terminal decision rule DT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50, $\delta_0 = .4$ with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKR	10	4.1662	3.7840	3.3167	2.8526
	30	11.9640	9.7450	7.3823	4.8030
	50	19.4748	15.1227	11.2932	6.5640
BKR*	10	5.4883	5.0599	5.2064	5.4941
	30	11.9576	12.1286	13.5288	15.6853
	50	17.8000	18.3092	21.2853	25.7188
PWR	10	4.2553	3.8446	3.3897	2.8975
	30	12.0644	9.6805	7.5570	5.2853
	50	19.6469	15.2241	11.4261	7.4238
PLR	10	5.4459	5.0065	5.0874	5.3544
	30	11.6540	11.8502	13.2081	15.3389
	50	17.4052	18.0040	20.8792	25.1340
PCWR	10	4.2145	3.6668	3.2382	3.2422
	30	11.6410	9.0828	7.2883	6.3885
	50	18.6903	13.9212	10.8003	8.8995

Table (7.31)

$E(N^*_{(1)})$ of the sampling rules BKR, BKR*, PWR, PLR and PCWR under the stopping rule DS and the terminal decision rule DT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50, $\delta_0 = .4$ with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKR	10	4.1953	3.9708	3.6409	3.1444
	30	12.0582	10.4978	9.0967	6.5759
	50	19.6992	16.7188	14.4777	9.9195
BKR*	10	5.7979	5.4167	5.3758	5.5267
	30	13.8478	14.0285	14.4010	15.7975
	50	21.3265	21.7031	22.8056	25.8901
PWR	10	4.2856	4.0326	3.7189	3.1603
	30	12.1635	10.5733	9.4325	6.9427
	50	19.8654	16.9295	14.4689	10.4189
PLR	10	5.7462	5.3367	5.2861	5.3754
	30	13.6188	13.6179	14.1679	15.4616
	50	20.7110	21.4653	22.4885	23.3419
PCWR	10	4.2500	4.0365	3.7924	3.6070
	30	11.8255	10.6032	9.7538	8.0031
	50	19.1097	16.5507	15.4553	11.4363

Table (7.32)

E(R) of the sampling rules BKR, BKR*, PWR, PLR and PCWR under the stopping rule DS and the terminal decision rule DT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50, $\delta_0 = .4$ with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKR	10	2.0634	3.7661	5.2363	7.0738
	30	5.9104	9.7801	12.2611	14.7666
	50	9.5949	15.1361	18.7924	21.2840
BKR*	10	1.3456	3.1749	5.2604	7.5801
	30	2.5556	7.3384	13.4986	21.6643
	50	3.6582	11.0388	21.1713	35.6539
PWR	10	2.0332	3.7089	5.2128	7.1023
	30	5.8222	9.4914	12.0547	14.8468
	50	9.4665	14.9439	18.5022	21.9630
PLR	10	1.3571	3.1852	5.2584	7.6078
	30	2.6587	7.3525	13.4534	21.6145
	50	3.7966	11.0903	21.0562	35.2608
PCWR	10	1.9400	3.3256	4.5709	6.4484
	30	5.5160	8.6542	11.0742	16.0224
	50	8.9051	13.4551	16.9090	24.5470

Table (7.33)

$E(R^*)$ of the sampling rules BKR, BKR*, PWR, PLR and PCWR under the stopping rule DS and the terminal decision rule DT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50, $\delta_0 = .4$ with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
BKR	10	2.1382	4.1309	6.0545	8.0650
	30	6.5413	12.7876	18.9447	25.3263
	50	11.0023	21.5104	31.8772	42.6613
BKR*	10	1.5796	3.7711	5.8086	7.8635
	30	5.9679	12.0332	17.9636	23.7604
	50	10.4645	20.5029	30.2378	39.7596
PWR	10	2.1258	4.1149	6.0492	8.0698
	30	6.5279	12.7960	18.8721	25.2744
	50	10.9493	21.4780	31.8687	42.5964
PLR	10	1.5980	3.7893	5.8023	7.8992
	30	6.0176	12.1175	18.0018	23.8520
	50	10.5821	20.5487	30.2999	39.8607
PCWR	10	2.1273	4.0840	5.9338	8.1533
	30	6.5810	12.7699	18.7247	25.2828
	50	11.0898	21.5637	31.5793	42.5258

Table (7.34)

Performance characteristics of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS and the terminal decision rule BKT for generated p_1, p_2, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	Performance characteristics					
		P(CS)	E(M)	$E(N_{(1)})$	E(R)	$E(N^*_{(1)})$	$E(R^*)$
A_2	10	.8176	6.8840	4.3416	3.1899	4.7725	4.5723
	30	.8908	21.9560	14.1543	9.8248	14.6052	14.5525
	50	.9108	37.0478	23.9104	16.4615	24.3853	24.5415
A_5	10	.8169	6.8472	2.5208	3.7164	2.9530	5.4612
	30	.8909	21.8943	7.7513	12.1098	8.2128	17.6909
	50	.9144	37.0694	13.1571	20.5941	13.6059	29.8787
GS	10	.8089	7.6700	3.8350	3.8606	4.0890	5.0207
	30	.8860	22.9650	11.4825	11.4719	11.7829	16.0145
	50	.9156	38.3164	19.1582	19.1455	19.4548	27.0810
PWR	10	.8250	6.8017	2.5460	3.6748	2.9682	5.4362
	30	.8816	21.7637	7.6444	11.9621	8.1485	17.6595
	50	.9123	36.6640	12.5696	20.3499	13.0665	30.0404
PLR	10	.8204	6.8339	4.2496	3.1776	4.6828	4.5847
	30	.8916	21.7715	14.1281	9.8245	14.6111	14.6485
	50	.9104	36.7810	24.0731	16.7250	24.5888	24.8250
PCWR	10	.8213	7.2983	3.0472	3.8437	3.3267	5.3094
	30	.8901	22.2366	8.1342	12.1407	8.5136	17.5093
	50	.9120	37.2137	13.1732	20.5515	13.6022	29.8107

Table (7.35)

P(CS) of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	.7807	.7354	.7386	.7769
	30	.9232	.8660	.8681	.9212
	50	.9628	.9216	.9258	.9632
A_5	10	.7725	.7443	.7363	.7839
	30	.9182	.8746	.8663	.9231
	50	.9645	.9215	.9217	.9652
GS	10	.6863	.6747	.6785	.6848
	30	.8875	.8465	.8485	.8950
	50	.9535	.9116	.9176	.9574
PWR	10	.7750	.7394	.7337	.7754
	30	.9134	.8681	.8746	.9185
	50	.9628	.9234	.9308	.9626
PLR	10	.7848	.7346	.7364	.7812
	30	.9194	.8670	.8668	.9176
	50	.9624	.9265	.9225	.9637
PCWR	10	.6987	.6977	.7120	.7179
	30	.9015	.8544	.8580	.9136
	50	.9591	.9187	.9223	.9619

Table (7.36)

E(M) of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	6.2179	7.0318	7.6694	8.3041
	30	20.9036	23.4319	24.9190	26.1151
	50	35.6814	39.9975	42.1325	43.8511
A_5	10	8.2848	7.6667	7.0054	6.1978
	30	26.1042	24.8753	23.4646	20.9171
	50	43.8118	42.1316	39.9328	35.7631
GS	10	8.2160	7.9410	7.9588	8.2362
	30	25.1110	24.8350	24.7910	25.0788
	50	41.9378	41.7478	41.6538	41.8774
PWR	10	8.1640	7.6411	7.0307	6.1613
	30	26.1619	25.0386	23.2497	20.1581
	50	44.0529	42.4094	39.5799	34.0339
PLR	10	6.1781	7.0405	7.6349	8.1542
	30	20.1470	23.3083	24.9828	26.1322
	50	33.9044	39.6581	42.3479	44.0349
PCWR	10	8.5866	7.9254	7.4138	7.2591
	30	26.5296	25.3487	23.6649	20.9876
	50	44.4285	42.7334	39.9341	34.6857

Table (7.37)

$E(N_{(1)})$ of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	3.9078	4.1161	4.3567	4.5978
	30	13.8060	13.8969	14.0740	14.4938
	50	23.4927	23.5175	23.6931	24.2643
A_5	10	3.6873	3.3162	2.8910	2.2681
	30	11.6380	10.8036	9.6103	7.1392
	50	19.5394	18.4677	16.4553	12.3140
GS	10	4.1080	3.9705	3.9794	4.1181
	30	12.5555	12.4175	12.3955	12.5394
	50	20.9689	20.8739	20.8269	20.9387
PWR	10	3.7452	3.3762	2.9478	2.2410
	30	11.7277	10.7812	9.1186	6.1062
	50	19.5844	18.0857	15.3455	9.7610
PLR	10	3.9581	4.0927	4.2620	4.4107
	30	14.0908	14.1088	14.2678	14.4480
	50	24.2750	24.2440	24.3446	24.4669
PCWR	10	3.9856	3.6235	3.3118	3.1516
	30	11.8794	11.0100	9.5833	7.0646
	50	19.7458	18.3284	15.7388	10.6352

Table (7.38)

$E(N^*_{(1)})$ of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	4.6965	4.7882	4.8472	4.8814
	30	14.2251	14.3944	14.4616	14.6819
	50	23.7171	23.8505	23.9563	24.3855
A_5	10	3.9867	3.8004	3.5689	3.0371
	30	11.8284	11.1702	10.1190	7.5762
	50	19.6478	18.7565	16.7859	12.5334
GS	10	4.3374	4.3813	4.3782	4.3453
	30	12.7089	12.7715	12.7309	12.6922
	50	21.0451	21.1211	21.0619	21.0095
PWR	10	4.0534	3.8727	3.6248	3.0417
	30	11.8802	11.1506	9.6103	6.6394
	50	19.6552	18.3402	15.6801	10.0253
PLR	10	4.7153	4.7593	4.7643	4.7121
	30	14.5901	14.6342	14.6252	14.5878
	50	24.5429	24.5823	24.5849	24.5345
PCWR	10	4.1662	4.0646	3.8670	3.5662
	30	11.9810	11.3458	10.0607	7.4903
	50	19.7938	18.5539	16.0608	10.8913

Table (7.39)

E(R) of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	1.0907	2.7150	4.5142	6.5644
	30	3.5103	8.9184	14.6262	20.6133
	50	6.0003	15.2783	24.7537	34.6251
A_5	10	1.7625	3.1940	4.3283	5.1217
	30	5.5164	10.2822	14.5008	17.3880
	50	9.2330	17.3641	24.6763	29.7192
GS	10	1.6536	3.2131	4.7850	6.5927
	30	4.9880	9.9343	14.8824	20.0619
	50	8.3524	16.7178	25.0255	33.5091
PWR	10	1.7040	3.1667	4.3482	5.1081
	30	5.4890	10.3638	14.4509	16.9288
	50	9.2722	17.5836	24.6577	28.6967
PLR	10	1.0827	2.7227	4.5167	6.4495
	30	3.2069	8.8250	14.6240	20.6250
	50	5.2953	14.9995	24.7873	34.7553
PCWR	10	1.7827	3.2714	4.5464	5.9135
	30	5.5760	10.4548	14.6706	17.4869
	50	9.3433	17.7026	24.8392	29.1045

Table (7.40)

$E(R^*)$ of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	1.6363	3.7267	5.6333	7.2597
	30	5.6814	11.8121	17.6433	23.4763
	50	9.7902	19.9455	29.7649	39.6361
A_5	10	1.7955	3.9300	5.8366	7.7917
	30	6.1519	12.4839	18.5449	24.8547
	50	10.6061	20.9641	31.2120	41.9164
GS	10	1.8198	4.0737	5.8056	7.6370
	30	6.0036	12.3960	18.0993	23.9667
	50	10.3649	20.7651	30.4556	40.2306
PWR	10	1.7523	3.9166	5.8535	7.8191
	30	6.1247	12.4817	18.6403	25.0959
	50	10.5860	21.0530	31.4335	42.4439
PLR	10	1.6062	3.7225	5.6510	7.2938
	30	5.6025	11.7780	17.6050	23.4848
	50	9.6162	19.8374	29.6515	39.6110
PCWR	10	1.8376	3.9271	5.8479	7.8009
	30	6.1328	12.4400	18.5821	24.9321
	50	10.5497	21.0230	31.3804	42.2710

Table (7.41)

Performance characteristics of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_1 and the terminal decision rule BKT for generated p_1, p_2, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	Performance characteristics					
		P(CS)	E(M)	$E(N_{(1)})$	E(R)	$E(N^*_{(1)})$	$E(R^*)$
A_2	10	.6980	2.9087	1.7641	1.4138	3.8668	5.0441
	30	.7700	16.5141	10.3935	7.8826	13.0314	15.4964
	50	.8059	29.3998	18.5209	13.9109	22.1522	25.8776
A_5	10	.7098	3.2679	1.2506	1.8322	3.2040	5.3546
	30	.8131	9.7329	3.0213	5.5508	6.5968	18.4172
	50	.8362	15.5405	4.3116	8.9448	9.5745	31.2505
GS	10	.7119	4.7972	2.3986	2.5505	3.8564	5.0249
	30	.8633	14.3814	7.1907	7.7702	8.8257	17.1643
	50	.8956	23.7950	11.8975	13.0490	13.8183	29.2413
PWR	10	.7079	3.2756	1.2458	1.8461	3.2053	5.3348
	30	.8547	10.4032	3.6271	5.9086	6.1946	18.3147
	50	.8861	17.4428	6.1020	10.0509	9.1637	31.1363
PLR	10	.7009	2.8531	1.7393	1.3817	3.8466	5.0596
	30	.8658	15.0740	9.5719	7.5096	11.1717	16.2209
	50	.9014	28.2384	18.0382	13.9331	19.5162	26.9169
PCWR	10	.7167	4.0910	1.8295	2.2660	3.3936	5.2434
	30	.8642	11.5399	4.4979	6.5042	6.5675	18.1465
	50	.8959	18.5143	6.9471	10.6396	9.4844	31.1900

Table (7.42)

P(CS) of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_1 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	.6033	.6201	.6319	.6321
	30	.8316	.7370	.6591	.5233
	50	.8759	.7472	.7088	.6112
A_5	10	.6074	.6216	.6443	.6684
	30	.8185	.7586	.7283	.7134
	50	.8719	.7860	.7354	.7055
GS	10	.2835	.4303	.5339	.5550
	30	.7959	.8000	.8272	.8706
	50	.8887	.8849	.9078	.9468
PWR	10	.6016	.6221	.6411	.6640
	30	.8288	.7933	.8036	.8477
	50	.8892	.8667	.8896	.9253
PLR	10	.6054	.6150	.6306	.6263
	30	.8391	.8171	.8354	.8970
	50	.9087	.8999	.9082	.9563
PCWR	10	.2890	.4296	.5249	.5489
	30	.7987	.7905	.8103	.8759
	50	.8904	.8737	.8873	.9399

Table (7.43)

E(M) of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_1 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	2.2717	2.7123	3.2269	3.6446
	30	12.7206	17.2489	20.0489	21.5203
	50	23.7377	31.9342	35.4545	37.4976
A_5	10	3.2030	3.3729	3.4234	3.2845
	30	11.9881	11.0477	9.9675	8.8902
	50	19.1272	17.6241	16.0938	14.4194
GS	10	4.1160	4.4886	5.0646	6.0828
	30	11.9044	14.4578	17.8276	21.1074
	50	18.2332	24.8336	31.5806	36.8056
PWR	10	3.1950	3.3723	3.4278	3.2949
	30	10.8435	11.8895	12.3317	11.5550
	50	17.0712	20.5810	22.3332	20.7994
PLR	10	2.2612	2.6522	3.0826	3.5477
	30	10.0610	15.2734	20.0496	22.9877
	50	18.4945	31.1441	37.8663	41.0444
PCWR	10	3.7429	3.8387	4.3123	5.3809
	30	11.6883	12.7380	13.6228	14.0761
	50	17.8500	21.7697	23.9580	23.3449

Table (7.44)

$E(N_{(1)})$ of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_1 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	1.2876	1.5223	1.8106	2.0577
	30	7.4672	9.9466	11.5836	13.1340
	50	13.6483	18.4499	20.4450	22.8401
A_5	10	1.4208	1.4698	1.4438	1.2961
	30	4.3211	4.0523	3.5854	3.1956
	50	5.9877	5.9497	5.5277	5.2323
GS	10	2.0580	2.2443	2.5323	3.0414
	30	5.9522	7.2289	8.9138	10.5537
	50	9.1166	12.4168	15.7903	18.4028
PWR	10	1.4237	1.4638	1.4532	1.3177
	30	4.6835	4.9846	4.7975	3.6791
	50	7.4186	8.6358	8.5468	6.1157
PLR	10	1.2871	1.4854	1.7272	1.9794
	30	6.8326	9.3215	11.6655	13.0089
	50	13.0118	19.2441	22.0748	23.1807
PCWR	10	1.7638	1.7987	2.0100	2.4893
	30	5.2539	5.5890	5.6797	5.2099
	50	7.9615	9.3913	9.6401	7.7711

Table (7.45)

$E(N^*_{(1)})$ of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_1 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	4.3902	4.3202	4.2594	4.2508
	30	9.9894	13.0012	15.2354	18.1501
	50	16.2306	23.1878	25.7583	29.9896
A_3	10	4.1128	3.9824	3.7823	3.5378
	30	7.3741	8.4433	8.9298	9.0867
	50	9.6070	12.5787	14.2943	15.2720
GS	10	6.2842	5.3887	4.7565	4.4186
	30	9.2630	10.0145	10.5046	11.0643
	50	12.3666	14.8682	16.7971	18.6262
PWR	10	4.1561	3.9793	3.8128	3.5915
	20	7.7758	8.5391	8.0948	6.4094
	50	10.7959	12.2389	11.3085	8.1388
PLR	10	4.3802	4.3400	4.2674	4.2922
	30	9.9411	11.7851	12.9097	13.5113
	50	15.6937	20.6514	22.7621	23.3918
PCWR	10	6.0363	5.1634	4.5072	4.0399
	30	8.4965	8.9142	8.4701	6.8750
	50	11.1224	12.5399	12.0624	8.9972

Table (7.46)

E(R) of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_1 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	.4339	1.0694	1.9094	2.8663
	30	2.3389	6.6207	11.6983	16.7384
	50	4.3860	12.2728	20.7188	29.2050
A_5	10	.6916	1.4196	2.1220	2.7094
	30	2.7251	4.7273	6.2697	7.3780
	50	4.5190	7.6416	10.1575	11.9472
GS	10	.8363	1.8265	3.0504	4.8738
	30	2.3831	5.7972	10.6949	16.8749
	50	3.6236	9.9340	18.9747	29.4449
PWR	10	.6826	1.4191	2.1305	2.7107
	30	2.3246	4.9706	7.6890	9.6701
	50	3.6315	8.5501	13.9017	17.4935
PLR	10	.4298	1.0390	1.8157	2.8024
	30	1.6576	5.7634	11.6886	18.0889
	50	2.9182	11.7405	22.0960	32.3194
PCWR	10	.7834	1.5845	2.6428	4.3433
	30	1.4613	5.2496	8.4177	11.6384
	50	3.7523	8.9946	14.8521	19.4560

Table (7.47)

$E(R^*)$ of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_1 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	1.6181	3.8935	5.7305	7.7873
	30	6.4879	12.0741	17.5285	22.8980
	50	11.2993	20.0442	29.4417	38.5740
A_5	10	1.7909	3.8988	5.5948	7.8027
	30	7.0772	13.0439	18.8623	24.5620
	50	12.5930	22.2127	31.7547	41.4531
GS	10	1.1140	3.5244	5.8667	7.6479
	30	6.7961	12.9401	18.5836	24.2275
	50	12.1191	21.9814	31.3271	40.6820
PWR	10	1.7742	3.8995	5.5974	7.7851
	30	7.0345	13.0281	19.0415	25.1093
	50	12.3957	22.2694	32.3157	42.8175
PLR	10	1.6197	3.8781	5.7123	7.7961
	30	6.5192	12.3660	17.9717	23.8209
	50	11.4042	20.6172	30.0160	39.8331
PCWR	10	1.3422	3.4407	5.6328	7.7423
	30	6.8067	12.9228	18.9254	25.0362
	50	12.2647	22.2059	32.1622	42.6194

Table (7.48)

Performance characteristics of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_2 and the terminal decision rule BKT for generated p_1, p_2, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	Performance characteristics					
		P(CS)	E(M)	$E(N_{(1)})$	E(R)	$E(N^*_{(1)})$	$E(R^*)$
A_2	10	.4160	8.5845	6.0974	3.4015	6.9376	3.6219
	30	.2473	24.8332	19.5765	8.8238	23.4423	10.7891
	50	.2084	40.5503	32.9386	13.7858	40.2872	17.7307
A_5	10	.7840	6.8159	2.2357	3.7930	2.8135	5.5188
	30	.8246	19.6498	5.0163	11.4003	6.5465	18.2881
	50	.8422	32.2621	7.2137	19.0171	9.5807	31.2245
GS	10	.8108	9.0436	4.5218	4.5074	4.5730	4.9087
	30	.8883	26.2788	13.1394	13.1561	13.2502	15.4515
	50	.9134	43.6148	21.8074	21.7692	21.9300	25.9836
PWR	10	.8160	7.2043	2.6030	3.9564	2.9893	5.4521
	30	.8891	22.2134	7.6374	12.2269	8.0606	17.6276
	50	.9123	37.2481	12.7111	20.5368	13.1736	29.8173
PLR	10	.5955	9.0000	5.9373	3.7791	6.3418	3.7791
	30	.5992	29.0000	19.7949	12.0416	20.1957	12.0416
	50	.5982	49.0000	33.6074	20.1301	34.0092	20.1301
PCWR	10	.8166	7.9120	3.1271	4.2914	3.3422	5.3358
	30	.8842	22.8770	8.2322	12.5715	8.5874	17.4756
	50	.9142	37.7805	13.1485	20.8622	13.5165	29.7493

Table (7.49)

P(CS) of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_2 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	.7388	.5167	.4104	.3489
	30	.4070	.2899	.2526	.1543
	50	.3244	.2620	.2040	.1020
A_5	10	.8448	.7343	.6955	.6899
	30	.8845	.7854	.7414	.7060
	50	.9254	.8131	.7528	.7049
GS	10	.6633	.6391	.6380	.6592
	30	.8785	.8256	.8349	.8838
	50	.9497	.9053	.9059	.9524
PWR	10	.8845	.7979	.7702	.7932
	30	.9479	.8872	.8919	.9204
	50	.9751	.9332	.9330	.9682
PLR	10	.8741	.8094	.7946	.7688
	30	.8705	.8155	.7948	.7795
	50	.8746	.8184	.7914	.7859
PCWR	10	.6847	.6851	.7023	.7241
	30	.8916	.8482	.8542	.9110
	50	.9601	.9166	.9194	.9607

Table (7.50)

E(M) of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_2 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	8.7680	8.4975	8.3078	8.3479
	30	26.7478	24.1393	22.5837	22.1097
	50	44.2669	39.2985	36.3371	34.8342
A_5	10	8.2826	7.4916	6.7045	5.8900
	30	24.6848	21.8534	19.2541	16.9059
	50	40.5942	35.8762	31.5849	27.7992
GS	10	9.4444	9.2828	9.3094	9.4688
	30	27.7532	27.5946	27.5790	27.7538
	50	46.0226	45.8914	45.8320	45.9714
PWR	10	8.5221	8.0130	7.4134	6.5180
	30	26.6159	25.4892	23.7750	20.6703
	50	44.5315	42.8975	40.0635	34.4717
PLR	10	9.0000	9.0000	9.0000	9.0000
	30	29.0000	29.0000	29.0000	29.0000
	50	49.0000	49.0000	49.0000	49.0000
PCWR	10	8.8972	8.4219	8.0067	7.9379
	30	26.8230	25.7511	24.1854	21.6535
	50	44.7055	43.1898	40.4807	35.4318

Table (7.51)

$E(N_{(1)})$ of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_2 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	5.7987	5.4838	5.2862	5.2902
	30	19.2201	17.3434	16.1205	16.2143
	50	32.0741	29.1478	27.5930	27.4758
A_5	10	3.1871	2.9133	2.6042	2.2135
	30	6.4986	6.7181	6.2902	5.7301
	50	8.1678	9.6841	9.6626	9.2105
GS	10	4.7222	4.6414	4.6547	4.7344
	30	13.8766	13.7973	13.7895	13.8769
	50	23.0113	22.9457	22.9160	22.9857
PWR	10	3.7978	3.4519	3.0170	2.2878
	30	11.7427	10.8056	9.2084	6.1643
	50	19.5877	18.1181	15.3646	9.7122
PLR	10	6.1447	5.6608	5.3734	5.0414
	30	21.1345	18.1426	17.0556	16.2805
	50	36.0913	30.6268	28.7175	27.4972
PCWR	10	4.0268	3.7344	3.4264	3.2385
	30	11.9088	11.0326	9.6105	7.0453
	50	10.7448	18.3850	15.7700	10.6732

Table (7.52)

$E(N^*_{(1)})$ of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_2 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	6.1087	6.2128	6.3083	6.4550
	30	20.6501	21.1308	21.5232	23.0350
	50	34.7130	36.2931	38.1515	41.2012
A_5	10	3.4055	3.5172	3.5434	3.4603
	30	6.8337	8.1393	8.7881	9.3511
	50	8.5243	11.7718	13.7007	15.2947
GS	10	4.7500	4.7350	4.7373	4.7618
	30	13.9150	13.9055	13.8993	13.9123
	50	23.0287	23.0295	22.9928	23.0045
PWR	10	3.9438	3.8136	3.5610	3.0140
	30	11.8300	11.0933	9.6197	6.6451
	50	19.6324	18.3242	15.6562	9.9492
PLR	10	6.2706	5.4702	5.1680	5.2726
	30	21.2640	17.9581	16.8504	16.5010
	50	36.2167	30.4452	28.5089	27.7113
PCWR	10	4.1564	4.0469	3.8461	3.5391
	30	11.9946	11.3141	10.0356	7.4371
	50	19.7841	18.5802	16.0593	10.9146

Table (7.53)

E(R) of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_2 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	1.4688	3.1696	4.7690	6.4558
	30	4.1570	8.5930	12.5693	16.6572
	50	6.8318	13.8099	19.9245	25.8727
A_5	10	1.8624	3.1786	4.1932	4.8718
	30	6.1205	9.5685	12.2289	14.0713
	50	10.5323	15.9891	20.1814	23.2036
GS	10	1.8976	3.7441	5.5952	7.5850
	30	5.5203	11.0442	16.5668	22.2090
	50	9.1809	18.3800	27.5356	36.7887
PWR	10	1.8056	3.3216	4.6033	5.4170
	30	5.6181	10.5615	14.8199	17.3661
	50	9.4198	17.8258	24.9927	29.0922
PLR	10	1.4715	3.4268	5.2928	7.0942
	30	4.4582	10.9090	16.9294	22.8380
	50	7.4797	18.4223	28.6251	38.6377
PCWR	10	1.8713	3.4837	4.9352	6.4983
	30	5.6507	10.6600	15.0195	18.0659
	50	9.4396	17.9261	25.2078	29.7809

Table (7.53)

E(R) of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_2 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	1.4688	3.1696	4.7690	6.4558
	30	4.1570	8.5930	12.5693	16.6572
	50	6.8318	13.8099	19.9245	25.8727
A_5	10	1.8624	3.1786	4.1932	4.8718
	30	6.1205	9.5685	12.2289	14.0713
	50	10.5323	15.9891	20.1814	23.2036
GS	10	1.8976	3.7441	5.5952	7.5850
	30	5.5203	11.0442	16.5668	22.2090
	50	9.1809	18.3800	27.5356	36.7887
PWR	10	1.8056	3.3216	4.6033	5.4170
	30	5.6181	10.5615	14.8199	17.3661
	50	9.4198	17.8258	24.9927	29.0922
PLR	10	1.4715	3.4268	5.2928	7.0942
	30	4.4582	10.9090	16.9294	22.8380
	50	7.4797	18.4223	28.6251	38.6377
PCWR	10	1.8713	3.4837	4.9352	6.4983
	30	5.6507	10.6600	15.0195	18.0659
	50	9.4396	17.9261	25.2078	29.7809

Table (7.54)

$E(R^*)$ of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_2 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	1.4688	3.3753	5.3917	7.1079
	30	4.5269	10.3434	16.4003	21.9847
	50	7.6293	17.3591	27.0669	36.2589
A_5	10	1.8930	3.9537	5.8862	7.7336
	30	7.1974	13.0634	18.8919	24.5693
	50	12.8496	22.3437	31.9058	41.3784
GS	10	1.8976	4.0559	5.9405	7.8506
	30	5.8410	12.2005	17.8674	12.7009
	50	9.9459	20.4004	30.0288	39.9014
PWR	10	1.8198	3.8908	5.8423	7.7866
	30	6.1271	12.4604	18.6426	25.0767
	50	10.5887	21.0609	31.4433	42.4529
PLR	10	1.4715	3.4268	5.2928	7.0942
	30	4.4582	10.9090	16.9294	22.8380
	50	7.4797	18.4223	28.6251	38.6377
PCWR	10	1.8790	3.9429	5.8852	7.7486
	30	6.1201	12.4588	18.5843	24.9260
	50	10.5558	21.0322	31.3797	42.2901

Table (7.54)

$E(R^*)$ of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_2 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	1.4688	3.3753	5.3917	7.1079
	30	4.5269	10.3434	16.4003	21.9847
	50	7.6293	17.3591	27.0669	36.2589
A_5	10	1.8930	3.9537	5.8862	7.7336
	30	7.1974	13.0634	18.8919	24.5693
	50	12.8496	22.3437	31.9058	41.3784
GS	10	1.8976	4.0559	5.9405	7.8506
	30	5.8410	12.2005	17.8674	12.7009
	50	9.9459	20.4004	30.0288	39.9014
PWR	10	1.8198	3.8908	5.8423	7.7866
	30	6.1271	12.4604	18.6426	25.0767
	50	10.5887	21.0609	31.4433	42.4529
PLR	10	1.4715	3.4268	5.2928	7.0942
	30	4.4582	10.9090	16.9294	22.8380
	50	7.4797	18.4223	28.6251	38.6377
PCWR	10	1.8790	3.9429	5.8852	7.7486
	30	6.1201	12.4588	18.5843	24.9260
	50	10.5558	21.0322	31.3797	42.2901

Table (7.55)

Performance characteristics of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_3 and the terminal decision rule BKT for generated p_1, p_2, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	Performance characteristics					
		P(CS)	E(M)	$E(N_{(1)})$	E(R)	$E(N^*_{(1)})$	$E(R^*)$
A_2	10	.6400	5.8130	3.8123	2.6292	5.0956	4.4782
	30	.3670	19.9626	15.1333	7.7207	20.3480	12.1483
	50	.2728	34.1180	27.2420	12.3732	37.3180	19.3029
A_5	10	.7526	4.5687	1.6259	2.5537	2.9255	5.4504
	30	.8122	12.4642	3.6236	7.1961	6.6804	18.2355
	50	.8344	20.0230	5.1522	11.7357	9.7409	31.1703
GS	10	.7708	5.9750			3.7571	5.0956
	30	.8791	18.9522			10.2443	16.8149
	50	.9074	30.8396			16.2538	28.6195
PWR	10	.7623	4.6844	1.7229	2.6065	2.9441	5.4485
	30	.8669	14.0018	4.8713	7.8963	6.5524	18.1398
	50	.9007	23.6099	8.2734	13.3977	10.0208	30.7709
PLR	10	.7199	5.9657	3.7753	2.7937	4.7179	4.6190
	30	.7648	23.2479	15.2686	10.7059	16.0887	14.0341
	50	.7864	41.7923	27.9524	18.6343	28.6757	22.7576
PCWR	10	.7668	5.6713	2.4060	3.1055	3.2299	5.3682
	30	.8690	15.0335	5.6654	8.5364	7.0745	18.0612
	50	.9006	24.5365	8.9700	14.0055	10.4927	30.9214

For GS, $E(N_{(1)}) = E(R) = \frac{E(M)}{2}$ since p_1, p_2 are generated from uniform distribution.

Table (7.56)

P(CS) of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_3 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	.7222	.6184	.5206	.4223
	30	.5793	.3344	.2583	.1580
	50	.4356	.2784	.2188	.1007
A_5	10	.7007	.6530	.6520	.6630
	30	.8251	.7573	.7269	.7080
	50	.8982	.7947	.7392	.7216
GS	10	.5651	.6186	.6464	.6413
	30	.8471	.8238	.8397	.8844
	50	.9298	.9038	.9125	.9531
PWR	10	.6990	.6577	.6787	.7154
	30	.8470	.8197	.8405	.8824
	50	.9290	.9008	.9128	.9542
PLR	10	.7428	.6675	.5914	.5346
	30	.8477	.7175	.5953	.4966
	50	.8816	.7219	.6053	.4956
PCWR	10	.5708	.6115	.6487	.6511
	30	.8436	.8167	.8311	.8964
	50	.9316	.9022	.9098	.9531

Table (7.57)

E(M) of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_3 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	5.1957	5.6159	6.1676	6.9365
	30	19.9937	20.2741	19.9521	20.6090
	50	36.5969	33.7277	32.0202	32.3795
A_5	10	5.0007	4.6964	4.5940	4.4098
	30	14.8704	13.6972	12.6220	11.6045
	50	23.7205	22.2256	20.4830	18.6540
GS	10	6.2164	5.7612	6.1304	7.2722
	30	16.6322	19.8680	23.0236	25.3686
	50	26.0056	34.1928	39.6876	43.1754
PWR	10	5.0285	4.8075	4.8050	4.6358
	30	14.9292	16.1286	16.5296	15.1249
	50	24.1514	28.3616	29.5556	26.5036
PLR	10	5.0226	5.6917	6.6316	7.5479
	30	18.5094	24.7749	27.2749	28.2263
	50	35.8546	45.2942	47.4489	48.2694
PCWR	10	6.0959	5.5727	5.7064	6.5876
	30	15.3364	16.8908	17.7731	17.1718
	50	24.8997	29.2591	30.6514	28.3393

Table (7.58)

$E(N_{(1)})$ of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_3 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	3.1092	3.3805	3.7264	4.2279
	30	13.7666	14.3710	14.1446	14.9393
	50	26.3408	24.8334	23.8732	25.3325
A_5	10	2.1288	1.9730	1.8519	1.7207
	30	4.9555	4.7681	4.3649	4.0307
	50	6.4206	6.8649	6.6980	6.1353
GS	10	3.1082	2.8806	3.0652	3.6361
	30	8.3161	9.9340	11.5118	12.6843
	50	13.0028	17.0964	19.8438	21.5877
PWR	10	2.2078	2.0666	1.9366	1.7033
	30	6.5211	6.7879	6.3559	4.5555
	50	10.5316	11.8832	11.2574	7.5599
PLR	10	3.1570	3.3764	3.8067	4.2194
	30	13.0702	15.3394	15.8564	15.8572
	50	26.0847	28.1186	27.6274	27.1107
PCWR	10	2.7938	2.5379	2.5518	2.8868
	30	6.8525	7.3254	7.2254	5.9679
	50	11.0616	12.5022	12.0959	8.9545

Table (7.59)

$E(N^*_{(1)})$ of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_3 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	4.3705	4.9293	5.4320	5.9318
	30	16.2107	20.0941	21.3700	22.9403
	50	30.9100	35.5454	37.5114	41.2690
A_5	10	3.5436	3.7732	3.7072	3.5918
	30	7.2959	8.5132	9.0114	9.2562
	50	8.7842	12.3252	14.1926	14.5003
GS	10	4.3402	4.2874	4.2284	4.1801
	30	9.8703	11.1804	12.0614	12.8231
	50	14.1958	17.8778	20.1698	21.6467
PWR	10	3.6035	3.7784	3.5596	3.2423
	30	8.4441	8.9864	8.2156	6.1429
	50	11.9483	13.4410	12.4938	8.3535
PLR	10	4.4080	4.7349	5.0175	5.1494
	30	14.4947	16.4359	16.7063	16.5860
	50	27.1634	29.0809	28.3970	27.8368
PCWR	10	4.0188	4.0223	3.8475	3.5429
	30	8.6427	9.3457	8.8052	6.9321
	50	12.3654	13.9353	13.1823	9.5300

Table (7.60)

E(R) of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_3 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	.9491	2.1607	3.5860	5.4013
	30	3.2429	7.2425	11.1149	15.5391
	50	5.6735	11.8830	17.6236	24.0785
A_5	10	1.0912	1.9802	2.8581	3.6423
	30	3.4666	5.8883	7.9813	9.6597
	50	5.8144	9.7231	12.9935	15.5683
GS	10	1.2572	2.3389	3.6880	5.8220
	30	3.3053	7.9445	13.8127	20.2953
	50	5.1698	13.6844	23.8424	34.5552
PWR	10	1.0826	2.0087	2.9860	3.8304
	30	3.1703	6.6933	10.2975	12.7067
	50	5.1207	11.8038	18.4483	22.3581
PLR	10	.8869	2.1849	3.9065	5.9453
	30	2.9156	9.3145	15.9392	22.2187
	50	5.5300	17.0274	27.7136	38.0320
PCWR	10	1.2860	2.3104	3.5062	5.3476
	30	3.2335	6.9656	10.9843	14.2626
	50	5.2345	12.1325	19.0492	23.7225

Table (7.61)

$E(R^*)$ of the sampling rules A_2, A_5, GS, PWR, PLR and PCWR under the stopping rule BKS_3 and the terminal decision rule BKT for fixed p_1, p_2 with $p_{[2]} - p_{[1]} = .2$, where N = 10, 30, 50 with uniform priors.

Sampling rule	N	$p_{[2]} - p_{[1]} = .2$			
		(.1, .3)	(.3, .5)	(.5, .7)	(.7, .9)
A_2	10	1.5472	3.6154	5.6060	7.2729
	30	5.2959	10.5908	15.4196	21.9691
	50	8.2224	17.4319	27.1685	36.2946
A_5	10	1.7821	3.9662	6.0261	7.9101
	30	7.0768	12.9861	18.7869	24.7563
	50	12.8556	22.2342	31.7729	41.5041
GS	10	1.6996	3.9564	5.9741	7.7895
	30	6.6149	12.6849	18.2494	23.9023
	50	11.7180	21.3918	30.6220	40.1277
PWR	10	1.7677	3.9588	6.0686	7.9477
	30	6.8977	12.9443	18.9018	25.2790
	50	12.1432	11.0409	32.0416	42.7823
PLR	10	1.5521	3.6806	5.7195	7.3741
	30	5.6207	11.3484	17.3337	22.9924
	50	9.0827	18.8237	29.0009	38.7626
PCWR	10	1.7296	3.7955	5.9340	7.8495
	30	6.7552	12.8105	18.7781	25.0180
	50	12.0843	21.9328	31.8915	42.5618

7.4 Discussion and conclusion

In this section we bring together all the schemes discussed in the previous sections and make some brief comparisons. In fact the motivation behind our analysis is to study the effect of different types of sampling rules with different stopping rules and to determine whether a useful sampling rule can be found that biases sampling towards the better of the two populations. The discussion in this section will cover two important criteria. First, P(CS)/E(M), the probability of correct selection per unit observation; second, a new one which can be described below.

Based on our numerical investigations, we found that $E(N_{(1)})$ is nearly linear function of E(M) for all schemes, that is

$$E(N_{(1)}) = \alpha + \beta^{*}E(M).$$

The slope β^{*}, which might be calculated as

$$\beta^{*} = \frac{\text{Max}(E(N_{(1)}) - \text{Min}(E(N_{(1)}))}{\text{Max}(E(M)) - \text{Min}(E(M))}$$

$$= \frac{[E(N_{(1)})|N = 50] - [E(N_{(1)})|N = 10]}{[E(M)|N = 50] - [E(M)|N = 10]},$$

can be used as a criteria to judge the performance of the schemes where the scheme improves as β^{*} decreases.

Table (7.62) presents results of P(CS)/E(M) for generated p_1 and p_2, N = 10(10)50 for total of 40 schemes which result from the combinations of the sampling rules A_5, BKR, PWR, PCWR, GS, A_2, BKR^{*}, PLR and the stopping rules DS, BKS, BKS_1, BKS_2,

Table (7.62)

The values of the ratio P(CS)/E(M) for various schemes with δ_0 = .4, generated p_1 and p_2 under uniform priors.

Stopping rule	Sampling rule	N				
		10	20	30	40	50
DS	A_5	.0933	.0556	.0404	.0324	.0270
	BKR	.0974	.0649	.0490	.0400	.0345
	PWR	.0985	.0640	.0402	.0399	.0346
	PCWR	.1085	.0685	.0525	.0414	.0361
	GS	.1001	.0633	.0478	.0388	.0330
	A_2	.0952	.0573	.0426	.0330	.0272
	BKR^*	.0995	.0646	.0502	.0405	.0345
	PLR	.0993	.0653	.0493	.0411	.0349
BKS	A_5	.1193	.0603	.0407	.0305	.0247
	BKR	.1183	.0607	.0408	.0310	.0247
	PWR	.1213	.0611	.0405	.0309	.0249
	PCWR	.1125	.0587	.0400	.0303	.0245
	GS	.1055	.0562	.0386	.0293	.0239
	A_2	.1188	.0600	.0406	.0307	.0246
	BKR^*	.1180	.0606	.0411	.0308	.0249
	PLR	.1200	.0607	.0410	.0309	.0248

Table (7.62) continued

The values of the ratio P(CS)/E(M) for various schemes with δ_0 = .4, generated p_1 and p_2 under uniform priors.

Stopping rule	Sampling rule	N				
		10	20	30	40	50
BKS_1	A_5	.2172	.1186	.0835	.0616	.0538
	BKR	.2172	.1220	.0832	.0645	.0522
	PWR	.2161	.1213	.0822	.0632	.0508
	PCWR	.1752	.1092	.0749	.0584	.0484
	GS	.1484	.0894	.0600	.0478	.0376
	A_2	.2400	.0807	.0466	.0346	.0274
	BKR^*	.2435	.0960	.0557	.0406	.0310
	PLR	.2457	.0983	.0574	.0421	.0319
BKS_2	A_5	.1150	.0617	.0420	.0323	.0261
	BKR	.1130	.0598	.0403	.0310	.0250
	PWR	.1133	.0582	.0400	.0303	.0245
	PCWR	.1032	.0559	.0387	.0298	.0242
	GS	.0897	.0490	.0338	.0258	.0209
	A_2	.0485	.0169	.0100	.0068	.0050
	BKR^*	.0579	.0241	.0148	.0107	.0083
	PLR	.0662	.0315	.0207	.0154	.0122

Table (7.62) continued

The values of the ratio P(CS)/E(M) for various schemes with δ_0 = .4, generated p_1 and p_2 under uniform priors.

Stopping rule	Sampling rule	N				
		10	20	30	40	50
BKS_3	A_5	.1647	.0938	.0652	.0506	.0417
	BKR	.1630	.0908	.0636	.0486	.0390
	PWR	.1627	.0906	.0619	.0471	.0381
	PCWR	.1352	.0823	.0578	.0447	.0367
	GS	.1290	.0670	.0464	.0365	.0294
	A_2	.1100	.0337	.0184	.0110	.0080
	BKR^*	.1162	.0455	.0277	.0195	.0151
	PLR	.1207	.0523	.0329	.0238	.0188

NB: For FSS, E(M) = N, then

N =	10	20	30	40	50
P(CS)/N =	.0815	.0431	.0296	.0225	.0181

BKS_3. The terminal decision rule DT is used with the stopping rule DS and BKT is used with other stopping rules. It is clear from this table that PCWR is superior to other sampling rules if used with DS; in addition, the table indicates that A_5 is superior to other sampling rules under the stopping rules BKS_1, BKS_2 and BKS_3. However, it follows from theorem in Jennison (1983) that all sampling rules achieve the same P(CS) if used with BKS and consequently achieve roughly the same P(CS)/E(M) if p_1 and p_2 are generated. Two typical curves of P(CS)/E(M) as a function of N are shown in Fig. 7.7.

In Table (7.63), we give some results of the slope β^* for δ_1 schemes where $\delta_0 = .4$ with geberated p_1 and p_2. Clearly, the scheme δ_1 (A_5) is superior to other δ_1 schemes if the performance is measured by β^*.

Table (7.64) contains also some results of β^* but for all sampling rules A_5, BKR, PWR, PCWR, GS, A_2, BKR^*, PLR in conjunction with the stopping rules DS, BKS, BKS_1, BKS_2, BKS_3. We note from this table that A_5 is the best sampling rule under the stopping rules DS, BKS_1, BKS_2 and BKS_3; whilst BKR is the best under BKS.

Fig. 7.8 shows that $E(N_{(1)})$ as a linear function of E(M) for two typical schemes.

On the basis of the results obtained from this study, we have drawn some general conclusions about the performance of the schemes.

(1) From a practical point of view, these suboptimal schemes are very easy to implement.

(2) As expected, the sampling rules A_2, BKR^*, PLR perform badly when $E(N_{(1)})$ is concerned. On the contrary, the sampling rules A_5, BKR, PWR, PCWR sample less frequently from the inferior population than from the superior population, making them particularly attractive for use in clinical trials.

(3) Although non of the sampling rules A_5, BKR, PWR, PCWR is uniformly better than others, A_5 seems to be a fairly good sampling rule to use when we have prior information on p_1 and p_2.

(4) The sampling rule BKR is the best under the stopping rule BKS.

(5) An increase in p_1 and p_2 always favours the stay on the winner sampling rules, that is A_5, BKR, PWR, PCWR; in addition, for small p_1 and p_2, the stay on the loser sampling rules, that is A_2, BKR^*, PLR might be led to better results than the stay on the winner and GS sampling rules.

Roughly speaking, if P(CS)/E(M) is of interest then the scheme (A_5, BKS_1) might be adopted. On the other hand, if we are interested in $E(N_{(1)})$ and E(M) then Table (7.64) which contains the values of β^* suggests that the scheme (A_5, BKS_2) should be preferred. If we take into account the ease of implementing the scheme, using the prior information and the overall optimality of performance then the scheme δ_2 should be chosen. However, the choice of the scheme depends very much on the aim of the experimenter.

Table (7.63)

The slope β^* for δ_1 schemes where δ_0 = .4, N = 10(10)50, generated p_1 and p_2 under uniform priors.

Scheme	δ_1 (A_1)	δ_1 (A_2)	δ_1 (A_3)	δ_1 (A_4)	δ_1 (A_5)	δ_1 (A_6)
β^*	.7573	.7519	.7210	.2510	.2418	.2546

Table (7.64)

The slope β^* for various schemes where δ_0 = .4, N = 10(10)50, generated p_1 and p_2 under uniform priors.

Sampling rule	Stopping rule				
	DS	BKS	BKS_1	BKS_2	BKS_3
A_5	.2418	.3519	.2494	.1957	.2282
BKR	.3825	.3286	.3204	.3255	.3261
PWR	.3905	.3357	.3428	.3364	.3461
PCWR	.3809	.3385	.3548	.3355	.3479
GS					
A_2	.7519	.6488	.6325	.8397	.8278
BKR^*	.6184	.6714	.6550	.7015	.6816
PLR	.6079	.6620	.6421	.6918	.6748

NB: β^* = .5 for the sampling rule GS under all stopping rules.

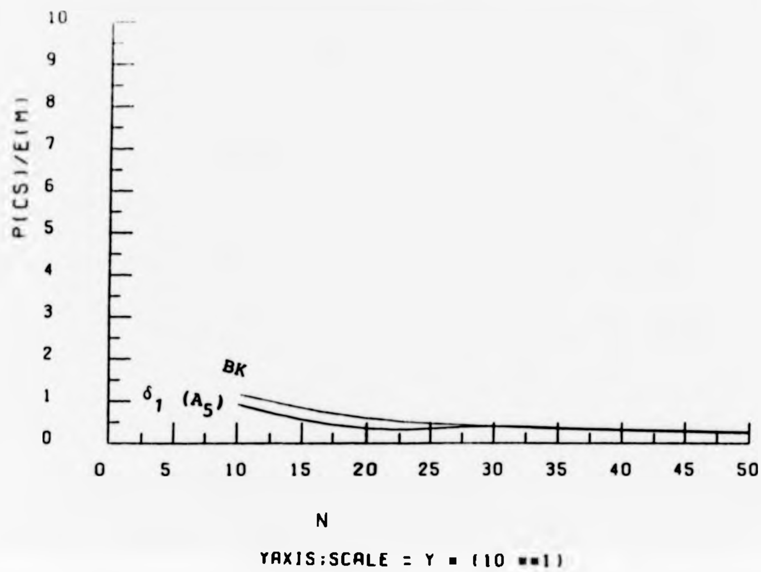

Fig. 7.7 The ratio P(CS)/E(M) as a function of N for the schemes (δ_1 (A_5), δ_0 = .4) and BK with uniform priors on p_1 and p_2.

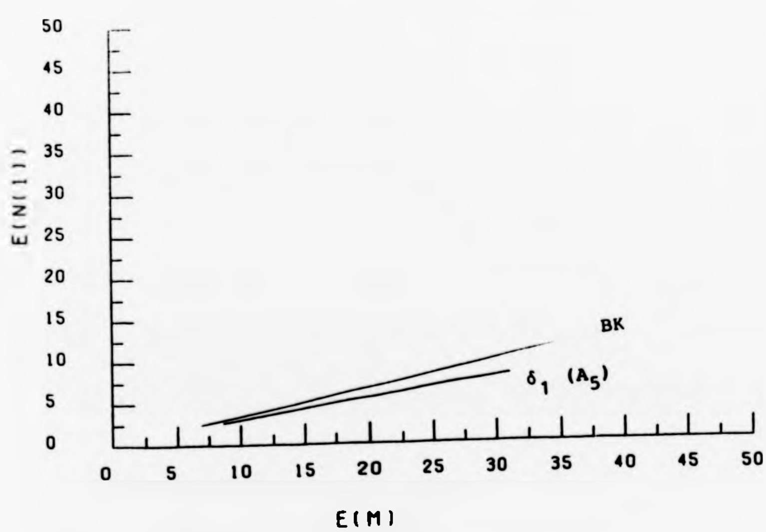

Fig. 7.8 $E(N_{(1)})$ as a function of E(M) for the schemes (δ_1 (A_5), δ_0 = .4) and BK where N = 10(10)50 with uniform priors on p_1 and p_2.

CHAPTER 8

BAYESIAN OPTIMAL AND SUBOPTIMAL DESIGNS FOR MINIMIZING THE EXPECTED NUMBER OF TRIALS ON THE INFERIOR POPULATION IN THE BINOMIAL SELECTION PROBLEM

8.0 Introduction and summary

Consider the problem of selecting the better of two Binomial populations as described earlier. The main objective in this chapter is to design some sequential selection schemes in such a way that a decision is reached with as small number of trials as possible on the inferior population, that is minimizing $E(N_{(1)})$. The problem with this objective has received considerable attention due to a possible application to the planning of sequential clinical trials where within this context the objective is to minimize the number of patients assigned to the poorer treatment.

Some authors have contributed to this probem; among them, Sobel and Weiss (1970), Hoel, Sobel and Weiss (1972, 1975a, 1975b) and Simon, Weiss and Hoel (1975). But these contributions used the indifference zone approach. Bechhofer and Kulkarni (1982) gave an upper bound for $E(N_{(1)})$ using the procedure BK described in chapter 7; later Jennison (1984) suggested a slight improvement on this upper bound and gave an approximate formula for $E(N_{(1)})$.

The optimal and suboptimal schemes we propose and discuss in the next sections are constructed using dynamic programming and a Bayesian approach where it is assumed that the success probabilities p_i $(i = 1, 2)$ are assigned independent Beta priors and where the total number of observations is N.

This chapter is organized into four sections.

In section 8.1 we derive the posterior probability that $p_1 > p_2$ and some of its properties.

Section 8.2 establishes the optimal and suboptimal schemes which are based on the dynamic programming and on the use of the posterior probability that $p_1 > p_2$.

A comparison of the results including those for the simple play-the-winner sampling rule are given in section 8.3.

8.1 Properties of the joint posterior distribution

As before, suppose that p_i is assigned a Beta prior distribution with parameters a_i and b_i, it is convenient to assume that these are integers, or $Be(a_i, b_i)$ with density proportional to

$$p_i^{a_i-1} \quad q_i^{b_i-a_i-1}, \qquad 1 \leqslant a_i \leqslant b_i - 1, \qquad q_i = 1 - p_i. \qquad (8.1.1)$$

Then at d_i trials with c_i successes, the posterior density of p_i is $Be(r_i, n_i)$ where $r_i = a_i + c_i$, $n_i = b_i + d_i$; it is assumed that p_i's are independent. The mean of the posterior distribution of p_i which may be used to estimate p_i is r_i/n_i.

The sequential designs in the next section require the derivation of the posterior probability that $p_1 > p_2$, denoted by $P(r_1, n_1, r_2, n_2)$, which is as follows. Let $\pi(p_i | r_i, n_i)$ be the posterior probability density function of p_i given r_i, n_i $(i = 1, 2)$, then

$$P(r_1, n_1, r_2, n_2) = \int_0^1 \int_0^{p_1} \pi(p_1, p_2 | r_1, n_1, r_2, n_2)\, dp_1\, dp_2 \tag{8.1.2}$$

$$= \frac{(n_1 - 1)!\,(n_2 - 1)!}{(r_1 - 1)!(n_1 - r_1 - 1)!(r_2 - 1)!(n_2 - r_2 - 1)!}$$

$$\int_0^1 \left[\int_0^{p_1} p_2^{\,r_2-1} (1 - p_2)^{n_2-r_2-1}\, dp_2\right]$$

$$p_1^{\,r_1-1} (1 - p_1)^{n_1-r_1-1}\, dp_1,$$

provided that $0 < r_i \leq n_i - 1$ $(i = 1, 2)$.

$$= \frac{(n_1 - 1)!\,(n_2 - 1)!}{(r_1 - 1)!\,(n_1 - r_1 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{1}{j!\,(n_2 - 1 - j)!}$$

$$\left[\int_0^1 p_1^{\,r_1-1+j} (1 - p_1)^{n_1+n_2-r_1-j-2}\, dp_1\right]$$

$$= \frac{(n_1 - 1)!\,(n_2 - 1)!}{(r_1 - 1)!\,(n_1 - r_1 - 1)!}$$

$$\sum_{j=r_2}^{n_2-1} \frac{\beta(r_1 + j,\, n_1 + n_2 - r_1 - j - 1)}{j!\,(n_2 - 1 - j)!}$$

$$= \sum_{j=r_2}^{n_2-1} G(j) \tag{8.1.3}$$

where

$$G(j) = \frac{(n_1 + n_2 - 1)\ \beta(r_1 + j + 1,\ n_1 + n_2 - r_1 - j - 1)}{j(r_1 + j)\ \beta(r_1,\ n_1 - r_1)\ \beta(j,\ n_2 - j)},$$

and $\beta(.,\ .)$ is a Beta function.

The form of the probability is complex and it is difficult to infer its behaviour on varying r_i, n_i, $i = 1, 2$. Nevertheless, it is possible to obtain results which give an idea of its properties and which are useful later in constructing the suboptimal stopping rule.

Define,

$$\Delta(r_1) = P(r_1 + 1,\ n_1,\ r_2,\ n_2) - P(r_1,\ n_1,\ r_2,\ n_2)$$

and

$$\Delta(n_1) = P(r_1,\ n_1 + 1,\ r_2,\ n_2) - P(r_1,\ n_1,\ r_2,\ n_2)$$

with $\Delta(r_2)$, $\Delta(n_2)$ similarly defined, then the following set of results hold.

<u>Theorem</u>

The following results hold at the point $(r_1,\ n_1,\ r_2,\ n_2)$

(a) If $r_1/(n_1 - 1) < r_2/(n_2 - 1)$ then $\Delta(r_1) > 0$

(b) If $r_1/n_1 < r_2/(n_2 - 1)$ then $\Delta(n_1) < 0$

(c) $\Delta(r_2) < 0$

(d) If $r_1/(n_1 - 1) < r_2/n_2$ then $\Delta(n_2) > 0$.

Proof

Denote the coefficient of G(j) in the following summations as C(j)

(a) $$\Delta(r_1) = \sum_{j=r_2}^{n_2-1} G(j)\left\{\frac{(r_1 + j)\ (n_1 - r_1 - 1)}{(n_1 + n_2 - r_1 - j - 2)\ r_1} - 1\right\}$$

$\Delta(r_1)$ will certainly be non-negative if $C(j) \geq 0\ \forall j$, which implies $C(r_2) \geq 0$, the result follows; note $C(n_2 - 1) \geq 0$.

(b) $$\Delta(n_1) = \sum_{j=r_2}^{n_2-1} G(j)\left\{\frac{(n_1 + n_2 - r_1 - j - 1)\ n_1}{(n_1 - r_1)\ (n_1 + n_2 - 1)} - 1\right\}$$

result follows from the condition $C(r_2) \leq 0$; note $C(n_2 - 1) < 0$.

(c) $$\Delta(r_2) = \sum_{j=r_2+1}^{n_2-1} G(j) - \sum_{j=r_2}^{n_2-1} G(j)$$

$= -G(r_2) < 0$ for all j.

(d) $$\Delta(n_2) = H(n_2) + \sum_{j=r_2}^{n_2-1} G(j)\left\{\frac{(n_1 + n_2 - r_1 - j - 1)\ n_2}{(n_1 + n_2 - 1)\ (n_2 - j)} - 1\right\}$$

where

$$H(n_2) = \frac{(n_1 - 1)!\ (n_2 + r_1 - 1)!}{(r_1 - 1)!\ (n_1 + n_2 - 1)!}$$

$\Delta(n_2)$ will certainly be positive if $C(r_2) \geq 0$, the result follows.

Stronger results may be possible; to illustrate the above, consider the "indifference" point (r, n, r, n) then $\Delta(r) \geq 0$ if $j \geq r$ for all j and the other Δ-inequalities all hold.

8.2 Sequential schemes

Let the total number of trials be N and suppose it is required to find a design which takes single observations from one of the two populations and after the result of each is known a decision is made on which population to sample next such that the expected number of observations on the inferior (smaller p_i) population is minimized. The first objective is to define the sampling rule, denoted by O-SR, which uses all the available observations.

At the point (r_1, n_1, r_2, n_2) in four dimensional integer space, let

$V(r_1, n_1, r_2, n_2)$ be the minimum expected number of trials on the inferior population in the remaining $(N - n_1 - n_2 - b_1 - b_2)$ trials,

$V_i(r_1, n_1, r_2, n_2)$ be the minimum number of trials if the next observation is from population i $(i = 1, 2)$.

The dynamic programming equations giving the partition of space into sample 1 and sample 2 points are:

$$V(r_1, n_1, r_2, n_2) = \min_{i=1,2} [V_i(r_1, n_1, r_2, n_2)]$$

where

$$V_1(r_1, n_1, r_2, n_2) = [1 - P(r_1, n_1, r_2, n_2)] + \frac{r_1}{n_1} V(r_1 + 1, n_1 + 1, r_2, n_2) + \left(1 - \frac{r_1}{n_1}\right) V(r_1, n_1 + 1, r_2, n_2)$$

with V_2 being defined in a similar manner.

The optimal decision at each point is now found by using the equations recursively from $(d_1 + d_2) = N$ where $V = 0$ to the origin $(d_1 + d_2) = 0$.

Note that at points where $(d_1 + d_2) = (N - 1)$ the optimal decision is to take an observation from population 1 if $P(r_1, n_1, r_2, n_2) > \frac{1}{2}$ and 2 if $P(r_1, n_1, r_2, n_2) < \frac{1}{2}$ and at random otherwise. This suggests a suboptimal sampling rule, denoted by SUB-SR, based on $P(r_1, n_1, r_2, n_2)$ only for all points. This is a one step ahead design since it proceeds as if sampling stops after the next observation thereby setting $V = 0$ at the transition points in the above equations; these designs have been discussed in other, related, contexts by Jones (1974, 1975). Using this suboptimal rule will naturally produce a design with a larger overall expected number of observations on the inferior population. The design may be used in a forward sense and does not need to be recomputed for different N.

The two sampling rules O-SR and SUB-SR would be more efficient if early stopping were allowed; hence these two sampling rules were investigated using the stopping rules DS, BKS, BKS_1, BKS_2 and BKS_3, which are described previously. Some numerical results are presented in the next section.

The dynamic programming equations giving the optimal sampling rule must be modified to allow for early stopping by setting V = 0 when the stopping rule is satisfied. This new design will not now be optimal with respect to both sampling and stopping rules.

The terminal decision rule DT (BKT) is used with the stopping rule (rules) DS (BKS and its modifications) respectively.

8.3 Results and discussion

In this section, our aim is to make numerical comparisons between the designs described in the previous section. For the sake of comparison between the Bayesian sampling rules O-SR and SUB-SR with the non-Bayesian sampling rule PWR, we assume that we are in a situation where little is known a priori about the values of p_1 and p_2. In other words, we assume that the information we have, concerning p_1 and p_2, comes primarily from the sample or equivalently $p_i \sim Be(1, 2)$, $i = 1, 2$.

Table (8.1) gives exact results for $E(N_{(1)})$ for the two sampling rules O-SR and SUB-SR and simulation results for PWR for various values of N. This table provides a direct comparison between all three sampling rules since no stopping rules are used. The differences between O-SR and SUB-SR are small for small N and increases as N increases with PWR performing rather badly. The ratio $E(N_{(1)})/N$ decreases with increasing N but remains roughly constant for PWR.

When the stopping rules are applied the inter-relation of the values of $E(N_{(1)})$ becomes difficult since the overall expected number of trials, $E(M)$, is not known for the designs.

Since a terminal decision is required, other characteristics such as the probability of making the correct decision (or correct selection) P(CS) would be of interest.

Table (8.2) illustrates the influence of the parameter δ_0 on $E(N_{(1)})$ when the stopping rule DS is used. It is clear that the smaller values of δ_0 result in lower values of $E(N_{(1)})$; however, the rate of increase in $E(N_{(1)})$ decreases as N increases. In fact there exists a value of δ_0 after which the values of $E(N_{(1)})$ remain constant.

Table (8.3) presents some values of $E(N_{(1)})$ for the three sampling rules used in conjunction with the five stopping rules; δ_0 was set to 0.3 in DS to reflect a balance between small $E(N_{(1)})$ for small δ_0 and large $E(N_{(1)})$ for large δ_0. It has to be remembered that if stopping is too early then there would be likely to be large error probabilities associated with the terminal decision. The same table clearly shows that the sampling rule O-SR is the best under all stopping rules; further, SUB-SR is uniformly better than PWR under all stopping rules except BKS, where PWR is better than SUB-SR for all N, and DS, where it is better than SUB-SR for relatively small N ($N \leq 40$).

The relative efficiency, defined by

$$\frac{[E(N_{(1)})\,|\,\text{O-SR}]}{[E(N_{(1)})\,|\,SR^*]} \times 100$$

where SR^* is any other sampling rule, is high for small sample sizes and decreases as N increases under all schemes. For example, the relative efficiency of the scheme (SUB-SR, DS) decreases from 93.6% to 61.6% as N increases from 10 to 100. With respect to the stopping rules, it is noted that DS (BKS_1) is the best under the sampling rules O-SR and PWR for large (small) sample sizes N. The stopping rule BKS_1 is the best under the sampling rule SUB-SR. BKS performs rather badly under all sampling rules.

In order to enable a direct comparison of the designs with the stopping rules to allow for varying the expected number of observations, E(M), some performance characteristics of PWR are presented in Table (8.4a). In Table (8.4b), $E(N_{(1)})$ for the optimal and suboptimal designs used without stopping rules where the sample size N is set to integer values greater than the five values of E(M) obtained by using the five stopping rules with PWR. This then forces the optimal and suboptimal designs to have the same E(M) as the PWR. These may now be compared directly with $E(N_{(1)})$ for PWR in Table (8.3). As a further comparison, the characteristic $E(N^*_{(1)})$ was computed for PWR, this is the expected number of trials on the inferior population if sampling is continued to N with the better population chosen at termination. Since $E(N^*_{(1)})$ is a function of both $E(N_{(1)})$ and P(CS) for PWR, it may be compared with the values of $E(N_{(1)})$ in Table (8.1). Using these, the suboptimal sampling rule SUB-SR shows a marked superiority over PWR.

The general conclusion that can be drawn is that the sampling rule SUB-SR performs well compared with O-SR with and without stopping rules; further, if we consider the ease of implementing the sampling rule then SUB-SR should be chosen. The stopping rule BKS_1 is uniformly better than other stopping rules if used in conjunction with the suboptimal sampling rule SUB-SR.

Table (8.1)

$E(N_{(1)})$ for the three sampling rules O-SR, SUB-SR and PWR where N = 10(10)100, all N observation are taken with uniform priors on p_1 and p_2.

N	Sampling rule		
	O-SR	SUB-SR	PWR
10	2.8155	2.9297	3.3199
20	4.5655	4.9870	6.4292
30	6.0020	6.8668	9.4895
40	7.2584	8.6646	12.5356
50	8.3936	10.4149	15.6444
60	9.4371	12.1346	18.6898
70	10.4099	13.8327	21.9253
80	11.3260	15.5143	24.8994
90	12.1942	17.1837	27.9795
100	13.0219	18.8431	31.1832

Table (8.2)

$E(N_{(1)})$ as a function of δ_0 where the stopping rule DS is used with the sampling rule O-SR with N = 50 and uniform priors on p_1 and p_2.

δ_0 =	0.0	0.1	0.2	0.3	0.4	0.5
$E(N_{(1)})$ =	0.5	0.5	1.8164	4.4749	6.7892	7.9534

δ_0 =	0.6	0.7	0.8	0.9	1.0
$E(N_{(1)})$ =	8.3823	8.3936	8.3936	8.3936	8.3936

Table (8.3)

$E(N_{(1)})$ for the combinations of the sampling rules O-SR, SUB-SR and PWR with the stopping rules DS, BKS, BKS_1, BKS_2 and BKS_3 where N = 10(10)100 and δ_0 = .3 under uniform priors on p_1 and p_2.

Sampling rule	N	Stopping rule				
		DS	BKS	BKS_1	BKS_2	BKS_3
O-SR	10	2.1855	2.4996	1.0750	2.2657	1.4884
	20	2.9921	4.8888	2.0502	3.6930	2.5839
	30	3.5743	7.3133	2.8391	4.8712	3.4716
	40	4.0591	9.7614	3.5924	5.9011	4.2348
	50	4.4749	12.2230	4.0998	6.8355	4.9282
	60	4.8368	14.6933	4.8053	7.6975	5.5654
	70	5.1388	17.1697	5.3292	8.5026	6.1566
	80	5.4201	19.6503	5.7477	9.2620	6.7157
	90	5.6845	22.1340	6.2416	9.9823	7.2481
	100	5.9351	24.6201	6.7592	10.6692	7.7558

Table (8.3) continued

$E(N_{(1)})$ for the combinations of the sampling rules O-SR, SUB-SR and PWR with the stopping rules DS, BKS, BKS_1, BKS_2 and BKS_3 where N = 10(10)100 and δ_0 = .3 under uniform priors on p_1 and p_2.

Sampling rule	N	Stopping rule				
		DS	BKS	BKS_1	BKS_2	BKS_3
SUB-SR	10	2.3346	2.6368	1.0750	2.2895	1.4988
	20	3.4364	5.3084	2.2499	3.8279	2.7698
	30	4.3502	8.0214	3.1739	5.1933	3.7689
	40	5.1842	10.7495	4.0952	6.4763	4.6621
	50	5.9739	13.4885	4.7007	7.7124	5.5006
	60	6.7359	16.2329	5.5945	8.9180	6.3019
	70	7.4786	18.9818	6.2696	10.1018	7.0791
	80	8.2068	21.7338	6.8556	11.2696	7.8379
	90	8.9243	24.4882	7.5091	12.4246	8.5826
	100	9.6332	27.2443	8.2818	13.5697	9.3137
PWR	10	2.2857	2.5460	1.2458	2.6030	1.7229
	20	3.4084	5.0748	2.3665	5.1690	3.2617
	30	4.3309	7.6444	3.6271	7.6374	4.8713
	40	5.0130	10.0953	4.7446	10.2230	6.5942
	50	5.9806	12.5696	6.1020	12.7111	8.2737
	60	6.9198	15.2663	7.5257	15.3595	9.8443
	70	7.4252	17.8076	8.8393	17.6125	11.5370
	80	8.2740	20.2190	10.0317	20.2022	13.2280
	90	9.0205	22.5721	11.4352	22.6370	15.0868
	100	10.3993	25.1126	12.8751	25.4541	16.4814

Table (8.4a)

Performance characteristics[+] of the designs where the sampling rule PWR is used with the stopping rules DS, BKS, BKS_1, BKS_2 and BKS_3 with N = 50 and uniform priors on p_1 and p_2.

Performance characteristics	Stopping rule				
	DS	BKS	BKS_1	BKS_2	BKS_3
P(CS)	.8509	.9123	.8861	.9123	.9007
E(M)	14.5525	36.6640	17.4428	37.2481	23.6099
$E(N^*_{(1)})$	11.0816	13.0665	9.1637	13.1736	10.0208

\+ Monte Carlo simulation results based on 10,000 runs.

Table (8.4b)

$E(N_{(1)})$ for the sampling rules O-SR and SUB-SR, no stopping rules, different values of N with uniform priors on p_1 and p_2.

Sampling rule	N				
	15	37	18	38	24
O-SR	3.7470	6.8960	4.2489	7.0181	5.1673
SUB-SR	3.9915	8.1309	4.5939	8.3095	5.7522

CHAPTER 9

BAYESIAN SEQUENTIAL SCHEMES FOR CHOOSING THE BEST MULTINOMIAL CELL

9.1 Introduction and review of literature

Consider a multinomial distribution with (k + 1) cells with unknown probabilities of an observation in the i^{th} cell p_i, $i = 1, 2, \ldots, k + 1$, where $p_{k+1} = \left(1 - \sum_{i=1}^{k} p_i\right)$. Denote the ordered values of the p_i by $p_{[1]} \leq p_{[2]} \leq \ldots \leq p_{[k+1]}$, the values of the p_i, and the pairing of the i^{th} cell with the $p_{[j]}$ $(1 \leq i, j \leq k + 1)$ are assumed to be completely unknown. Given a sequential sample of sample size N, it is required to find the cell with the largest probability, that is to select the cell associated with $p_{[k+1]}$, which is from this point termed the best cell.

Suppose that the prior information about the p_i may be adequately represented by a member of the natural conjugate Dirichlet family of distributions with parameters n_i', $i = 1, \ldots, k$ and m' with density proportional to

$$\prod_{i=1}^{k} p_i^{n_i'-1} \times p_{k+1}^{m' - \sum_{i=1}^{k} n_i' - 1}, \quad p_i \geq 0, \quad \sum_{i=1}^{k+1} p_i = 1,$$

$$n_i' \geq 1, \quad \sum_{i=1}^{k} n_i' \leq m' - 1. \tag{9.1.1}$$

After m trial with n_i observations in the i^{th} cell,

$i = 1, 2, \ldots, k$, the posterior distribution of the p_i is Dirichlet with parameters $(n_i' + n_i)$, $i = 1, 2, \ldots, k$ and $(m' + m)$ with mean

$$\hat{p}_i = \frac{n_i' + n_i}{m' + m}, \quad i = 1, 2, \ldots, k;$$

$n_i'' = n_i' + n_i$ will be termed the posterior frequency in the i^{th} cell. It is to be noted that if in (9.1.1) the n_i' are equal, then we have uniformity of prior knowledge for all p_i $(i = 1, 2, \ldots, k + 1)$, that is the prior expected value of each component probability would be $1/(k+1)$. The distribution which has the density (9.1.1) is called Dirichlet distribution and denoted by $D(n_1', n_2', \ldots, n_k'; m')$.

There are many practical situations where a solution to this problem is required. Applications to the identification of the most popular television program has been noted in Dudewicz (1976) and to varietal selection and voter preference in Gibbons et al (1977).

Many authors have considered the problem of selecting the largest cell probability, mostly under the indifference zone approach. According to this approach, the experimenter has to define a measure of distance Δ as the ratio of $p_{[k+1]}$ to $p_{[k]}$ and specifies two predetermined constants P^*, Δ^* such that $1/(k+1) < P^* < 1$ and $1 < \Delta^* < \infty$. The preference zone is defined as $\Delta \geq \Delta^*$ for some specified Δ^*. The experimenter would like to have a scheme for selecting the best cell which satisfies the probability requirements

$$P(CS) \geq P^* \quad \text{whenever} \quad \Delta \geq \Delta^*. \tag{9.1.2}$$

Bechhofer, Elmaghrabi and Morse (1959) have considered a single-sample scheme in which the cell with the largest count is selected as the best cell with ties broken by randomization. Cacoullos and Sobel (1966), Alam (1971) and Ramey and Alam (1979) proposed and investigated sequential schemes where observations are taken one at a time. In the sequential sampling of Cacoullos and Sobel, called inverse sampling, sampling is stopped when the count in any cell is N^* where N^* is a predetermined integer. The cell with count N^* is chosen as the best cell. In the second paper, sampling is stopped when the difference between the largest cell count and the second largest cell count is r^*, where r^* is a predetermined integer. The cell which has the highest count when sampling is stopped is selected as the best cell. In the third paper, sampling is stopped when the count in any cell is N^* or when the difference between the largest cell count and the second largest cell count is r^*, where r^* and N^* are predetermined positive integers and $r^* \leq N^*$.

Hwang (1982) suggested a multistage scheme for the problem. Single-stage schemes for selecting the worst cell (the cell with the smallest probability), using the indifference zone approach, were studied by Alam and Thompson (1972) and Gibbons et al. (1977). Subset selection schemes, where the aim is to select a non-empty subset of cells which contains the largest or smallest cell probability with a probability at least equal to a preassigned number, P^*, have been proposed by some authors such as Gupta and Nagel (1967) and Berger (1980).

Kulkarni (1981) proposed a closed sequential scheme without referring to the probability requirements given above. However, in many situations, we may have some information about the unknown parameters prior to the experimentation; this knowledge may come from past experiences. Therefore, it is worth considering the Bayesian approach to the problem. Ramey and Alam (1980) proposed a Bayesian sequential sampling procedure using a decision theoretic approach. This procedure is optimal in a sense that it minimizes the average risk with respect to certain prior distribution on the cell probabilities, namely the Dirichlet distribution. However, some computational and analytical complexities are associated with the implementation of such optimal scheme, particularly for large sample sizes and a large number of cells. Therefore, from practical point of view, it is preferable to find some schemes which are easy and simple to implement and which have some good characteristics.

In this chapter some Bayesian suboptimal selection schemes for the multinomial selection problem are proposed and investigated and then compared with those proposed by Kulkarni (1981). The investigation has been carried out using Monte Carlo (MC) simulation methods.

The contents of the chapter can be summarized as follows:

Section 9.2 presents methods to generate Dirichlet and Multinomial random variables which are used to carry out MC simulations.

Section 9.3 discusses some sampling methods and stopping rules which are used to construct some suboptimal schemes for selecting the best cell in Multinomial distribution.

Some numerical results, for Multinomial selection schemes with three and four cells, with a discussion are given in section 9.4.

Some remarks are given in section 9.5.

9.2 The generating of Dirichlet and Multinomial random variables

The generating of Dirichlet random variables is required in the next section where some MC simulations are carried out using generated probabilities from a Dirichlet distribution. There are two methods available for generating Dirichlet random variables; they are described as follows.

(a) Method ME_1:

In this method the distribution functions $F(p_1)$, $F(p_2|p_1)$, ..., $F(p_k|p_1, p_2, ..., p_{k-1})$ are found successively and then inverted to give the p_i's (i = 1, ..., k). This method is difficult to use except in certain special cases. For example, consider the trinomial distribution and suppose that the joint prior for p_1 and p_2 is uniform or Dirichlet (1, 1; 3) then the steps of the method are:

1. $$f(p_1, p_2) = \frac{\Gamma(m')}{\Gamma(n_1') \;\; \Gamma(n_2') \;\; \Gamma(n_3')} \; p_1^{n_1'-1}$$

$$p_2^{n_2'-1} \; (1 - p_1 - p_2)^{n_3'-1} \; , \quad m' = \sum_{i=1}^{3} n_i',$$

$$0 < p_1 < 1, \quad 0 < p_2 < 1 - p_1 \text{ (or } p_1 + p_2 < 1)$$

$$= 2. \qquad (9.2.1)$$

2. The marginal density of p_1 is

$$f_1(p_1) = \int_0^{1-p_1} f(p_1, p_2)\, dp_2 = 2\int_0^{1-p_1} dp_2 = 2(1 - p_1),$$

$$0 < p_1 < 1 \qquad (9.2.2)$$

(i.e. $f_1(p_1) \sim Be(1, 3)$).

3. The cumulative distribution function of p_1 is

$$F_1(p_1) = \int_0^{p_1} f_1(x)\, dx = 2\int_0^{p_1} (1 - x)\, dx = 2p_1 - p_1^2, \qquad (9.2.3)$$

$F_1(p_1)$ is uniformly distributed over (0, 1)
(i.e. $F(P_1) \sim U(0, 1)$).

4. The conditional density of $p_2|p_1$ is

$$f_2(p_2|p_1) = \frac{f(p_1, p_2)}{f_1(p_1)} = \frac{1}{1 - p_1}, \quad 0 < p_2 < 1 - p_1, \qquad (9.2.4)$$

(i.e. $f_2(p_2|p_1) \sim U(0, 1 - p_1)$.

5. The cumulative distribution function of p_2 given p_1 is

$$F_2(p_2|p_1) = \int_0^{p_2} f_2(x|p_1)\, dx = \int_0^{p_2} \frac{1}{1 - p_1}\, dx = \frac{p_2}{1 - p_1},$$

$$0 < p_2 < 1 - p_1, \qquad (9.2.5)$$

$F_2(p_2|p_1)$ is uniformly distributed over $(0, 1 - p_1)$.

6. Generate a uniform variate y_1 from $F_1(p_1)$, then

$2p_1 - p_1^2 = y_1$ giving

$p_1^2 - 2p_1 + y_1 = 0$ implies $p_1 = \dfrac{2 \pm \sqrt{4 - 4y_1}}{2} = 1 \pm \sqrt{1 - y_1}$,

hence $p_1 = 1 - \sqrt{1 - y_1}$ since $0 < p_1 < 1$.

7. Generate a uniform variate y_2 from $F_2(p_2|p_1)$, then

$p_2 = y_2(1 - p_1)$, where p_1 is that value obtained by (9.2.6).

(9.2.7)

The uniform random variates are generated using the subroutine Gϕ5CAF of NAG Library. The values of (P_1, p_2) computed in steps 6 and 7 can be considered as the realization of random variables possessing the Dirichlet distribution that is to be simulated.

(b) Method ME_2:

This method is essentially based on the connection between Dirichlet and Gamma distributions. The Gamma distribution with parameters $(\nu, 1)$ has the following probability density function

$$g(x) = \frac{x^{\nu-1} e^{-x}}{\Gamma(\nu)} \quad \text{for } \nu > 0,\ x < 0, \tag{9.2.8}$$

$$= 0 \quad \text{otherwise.}$$

Let $G_\nu(x)$ denotes the Gamma distribution function and for ease of notation it will be convenient to refer to $G_\nu(x)$ as $G(\nu)$.

The routine Gϕ5DGF of NAG Library is used to generate random variates from Gamma distribution with parameters

$(\nu, 1)$, where ν must take non-negative integer or half-integer values, and uses the following method:

For $\nu = 1$, we have the exponential case

$$G(1) = 1 - e^{-x}$$

which is uniformly distributed over (0, 1).
Hence for each observation of $U \sim U(0, 1)$ there exists exactly one value of X such that

$$U = 1 - e^{-X}. \tag{9.2.9}$$

The inversion method of Atkinson and Pearce (1976), when used to generate observations from the exponential distribution, yields

$$X = -\log(1 - U). \tag{9.2.10}$$

Some slight increase in computational efficiency can be gained by noting that if U is the uniform variable so is 1 - U and an equivalent form to (9.2.10) is

$$X = -\log U. \tag{9.2.11}$$

In other words, a random variable X having G(1) is generated as -log U.

Suppose that the random variable X_i ($i = 1, \ldots, n'$) has G(1), then the random variable

$$Z = \sum_{i=1}^{n'} X_i \quad \text{is generated as} \quad -\sum_{i=1}^{n'} \log U_i, \tag{9.2.12}$$

where $U_1, U_2, \ldots, U_{n'} \sim U(0, 1)$.

A method for generating Gamma distribution with non-integer shape parameter ν is described by Cheng (1979).

The following steps are used to generate the random variables $(p_1, p_2, \ldots, p_k)$ having Dirichlet distribution $D(n_1', n_2', \ldots, n_k'; m')$ to produce some of the simulation results given in section 9.4.

1. Generate the random variable Z_j having $G(n_j')$ $(j = 1, \ldots, k + 1)$, using the routine Gϕ5DGF of NAG Library with parameters n_j' and 1, where $n_{k+1}' = m' - \sum_{j=1}^{k} n_j'$.

2. From Wilks (1962, pp.178), we have the following connection between the Gamma random variables $Z_1, Z_2, \ldots, Z_{k+1}$ and the Dirichlet random variables $p_1, p_2, \ldots, p_k$.
"If Z_j has an independent Gamma distribution with parameters $(n_j', 1)$ $(j = 1, \ldots, k + 1)$ and

$$p_j = Z_j / \sum_{\ell=1}^{k+1} Z_\ell \tag{9.2.13}$$

then $(p_1, p_2, \ldots, p_k)$ have k-variate Dirichlet distribution $D(n_1', n_2', \ldots, n_k'; m')$ where $k \geqslant 2$."

Given the values of $\underset{\sim}{p} = (p_1, p_2, \ldots, p_{k+1})$, we are now in a position to generate observations from the multinomial distribution with $(k + 1)$ cells with p_i $(i = 1, 2, \ldots, k + 1)$ as the probability of an observation in the i^{th} cell. Let $\underset{\sim}{n} = (n_1, n_2, \ldots, n_{k+1})$ represents the observed frequencies in the $(k + 1)$ cells of the multinomial distribution. Then the problem is to generate a random vector $\underset{\sim}{n}$ where $\underset{\sim}{n} \sim \mu(m, \underset{\sim}{p})$ and

$\mu(m, \underset{\sim}{p})$ represents a multinomial conditional on the total number of observations $m = \sum_{i=1}^{k+1} n_i$ and the cell probabilities $\underset{\sim}{p}$.

An easy method is to generate one observation at a time using the probabilities $p_1, p_2, \ldots, p_{k+1}$. If a set of m independent observations is generated one at a time from the distribution $\mu(1, \underset{\sim}{p})$, the joint distribution of the m observations, represented by the vector of cell frequencies $\underset{\sim}{n}$, is $\mu(m, \underset{\sim}{p})$. Hence a straightforward method of obtaining $\underset{\sim}{n}$ is to use the uniform distribution to generate one observation at a time and accumulate the cell frequencies. The assignment of a success to one of the (k + 1) cells is as follows: Generate a uniform variate u from U(0, 1) and assign 1 to n_j if

$$\sum_{i=1}^{j} p_i - p_j < u < \sum_{i=1}^{j} p_i \quad \text{for all } j = 1, \ldots, k + 1 \text{ and } \sum_{i=1}^{k+1} p_i = 1.$$

9.3 Sampling methods and stopping rules

Three sampling methods are studied; these are:

1. MR_1: fully sequential where observations are taken sequentially one at a time until a terminal decision is reached,

2. $MR_2(g)$: group sequential where g observations are taken at a time; N will usually be a multiple of g,

3. MR_3: fixed sample size where all N observations are taken.

The first two methods are used in conjunction with the following four stopping rules:

1. $MS_1(\delta_0)$: stop sampling at the sample size m when

$$\hat{p}_{[k+1]} - \hat{p}_{[k]} \geq \delta_0, \qquad (9.3.1)$$

where δ_0 $(0 \leq \delta_0 \leq 1)$ is preassigned and $p_{[i]}$ is the i^{th} ordered posterior mean; this may be reformulated as

$$n''_{[k+1]} - n''_{[k]} \geq (m' + m)\,\delta_0, \qquad (9.3.2)$$

where $n''_{[i]}$ is the i^{th} ordered posterior frequency $(i = 1, \ldots, k + 1)$.

2. $MS_2(f^*)$: stop sampling when

$$n''_{[k+1]} - n''_{[k]} \geq f^*, \qquad (9.3.3)$$

where f^* is a preassigned positive integer.

3. MS_3: Kulkarni (1981), modified by using posterior frequencies. Stop sampling at the first sample size m if there exists a cell i such that

$$n''_i \geq n''_j + N - m \quad \text{for all } i, j, i \neq j \ (1 \leq i, j \leq k + 1) \qquad (9.3.4)$$

or

$$n''_{[k+1]} - n''_{[k]} \geq N - m. \qquad (9.3.5)$$

4. MS_4: a less conservative modification of MS_3 incorporating $\hat{p}_i$. Stop sampling after m observations if

$$n''_{[k+1]} - n''_{[k]} \geq (N - m)\,\hat{p}_{[k]} \tag{9.3.6}$$

where the term on the right-hand side represents the current posterior expected number of observations in the second largest frequency cell for the remaining N - m trials.

In all cases the best cell is chosen to be that with the largest posterior frequency with ties being broken randomly.

9.4 Results and Discussion

The criteria used to judge the performance of the rules are:

P(CS) = Probability of correct selection,

E(M) = Expected sample size,

P(M < N) = Probability of stopping before N,

P(CO) = Probability that the correct ordering of the cell probabilities is obtained.

The simulation computer program was written in FORTRAN and run on CYBER 205 at UMRCC. The program generates the necessary random numbers as input data for the simulation model and analyses the behaviour of the scheme. A listing of the program is available in the appendix (9.1).

The above criteria are calculated in each case from the results of a Monte Carlo simulation of 10,000 runs. As can be seen from Table (9.1), the vlaues of P(CS) and E(M) increase as the value of the parameter δ_0 increases but the rate of increase in these values decreases as δ_0 increases. There is also increase in P(CO) as δ_0 increases. The results for P(CS) and E(M) in Table (9.2) show the same behaviour with respect to f^*. Therefore the values of the parameters δ_0 and f^* in the stopping rules MS_1 (δ_0) and MS_2 (f^*) respectively were chosen by balancing P(CS) and E(M).

The fully sequential method MR_1 has smaller expected sample sizes than the group sequential MR_2 (5); this in turn generates a marginally smaller P(CS) for each stopping rule with MR_1. Naturally MR_2 (5) is easier to use since the stopping rule needs to be consulted less often. The stopping rule MS_2 (3) seems to dominate all others used in that it achieves an impressive level of correct decisions based on small sample sizes. The values of the ratio P(CS)/E(M), the probability of correct selection per unit actually observed, provide clear evidence of the domination of MS_2 (3) over others. For example, in Table (9.3) when N = 100 and the sampling method is MR_1, the values of this ratio are .0992, .0425, .0197 and .0120 under MS_2 (3), MS_1 (.3), MS_4 and MS_3 respectively. Using the same example but with the sampling method MR_2 (5), the values of the ratio are .0679, .0269, .0184 and .0093 under MS_2 (3), MS_1 (.3), MS_4 and MS_3 respectively.

The modified Kulkarni stopping rule MS_4 seems superior to the unmodified one MS_3.

Based on the values of P(CS)/E(M) calculated from Table (9.3), the performance of the various stopping rules can be ordered as follows:

$$MS_2\ (3) > MS_1\ (.3) > MS_4 > MS_3$$

under both sampling methods MR_1 and MR_2 (5), where (>) means better.

If the sampling method MR_3 is used, then its P(CS) increases slightly over others but that compensated for by using all N observations and consequently it is poorer than others if it is evaluated in terms of P(CS)/E(M).

In general the stopping rule MS_4 is the best as far as $P(M < N)$ is concerned under both sampling methods MR_1 and MR_2 (5). Under this criterion MS_1 (.3) and MS_2 (3) are conservative for very small p_i for both MR_1 and MR_2 (5). Surprisingly, the stopping rule MS_3 performs very well under MR_1 whilst it performs very badly under MR_2 (5).

It is obvious from Table (9.3) that P(CO) is an increasing function of N for all schemes. For small values of N, P(CO) $\underset{\sim}{\sim}$ ½ for all schemes and for large values of N the performance of the schemes can be ordered as

$MS_3 > MS_4 > MS_2$ (3) $> MS_1$ (.3)   under MR_1

and

$MS_3 > MS_4 > MS_1$ (.3) $> MS_2$ (3)   under MR_2 (5).

P(CO) achieves an impressive level over the a priori value of 1/6.

The effect of group size in the $MR_2(g)$ sampling methods was further investigated; some results for N = 100, generated p_1, p_2 with a uniform prior and g = 1, 5, 10, 25, 50 are presented in Table (9.4) for various stopping rules. Here there are great differences between the stopping rules in terms of E(M). The values of P(CS) for MS_3 and MS_4 take roughly the same values for all g nearly equal to that obtained with a fixed sample size of 100; therefore they may be compared in terms of E(M) with MS_3 requiring a much larger number of observations to termination than MS_4. The same can be said about MS_1 (.3) and MS_2 (3) as they achieve roughly the same P(CS) for all g; but they differ in E(M) where MS_3 (3) has much smaller E(M) than MS_1 (.3). However in terms of the values of the ratio P(CS)/E(M), the performance of the different stopping rules for different g and generated p_1, p_2 with uniform priors, can be ordered as follows

$$MS_2\ (3) > MS_1\ (.3) > MS_4 > MS_3.$$

As g increases, E(M) and hence P(CS) increases for all schemes. It is noticeable that

$$E(M|MR_2(g_2)) - E(M|MR_2(g_1)) \simeq g_2 - g_1, \qquad (g_1,\ g_2 \neq 1),$$

where $E(M|MR_2(g))$ is the expected sample size using sampling rule $MR_2(g)$ and stopping rule MS_2 (3).

Table (9.4) also demonstrates how $P(M < N)$ decreases rapidly as g increases for all stopping rules except for MS_2 (3) where the rate of decrease in $P(M < N)$ is small as g increases. In all schemes the values of $P(M < N)$ are high for small g. In terms of P(CO) values, given in the same table, the ordering of the performance is

$$MS_3 > MS_4 > MS_1\ (.3) > MS_2\ (3).$$

The results in Tables (9.1 - 9.4), obtained for generated

p_1 and p_2, give an idea of the relative merits of the schemes averaged over the prior but in practice their performance should be judged for p_i fixed (i = 1, 2), consequently a series of simulations were carried out for a range of values of p_i; some results are presented in Table (9.5). These results illustrate the errors which could occur if the sample sizes are too small which they tend to be in the adaptive stopping rules. Further, all the stopping rules would be rather conservative for very small p_i.

An example for four cell case is presented in Table (9.6) for p_i (i = 1, 2, 3) generated from a uniform prior. A comparison between the results in Tables (9.3) and (9.6) shows that as the number of cells increases, P(CS) decreases and E(M) increases while the values of P(CS)/E(M) are roughly constant for both cases.

A further example for four cell case is given in Table (9.7) for p_i (i = 1, 2, 3) generated from a non-uniform prior.

The superiority of the stopping rule MS_2 (3) over other stopping rules seems relatively unaffected by the change in the number of cells while the performance of MS_4 seems to improve.

Table (9.8) shows the degree of accuracy of our simulation results compared with the exact results obtained by Kulkarni (1982), in terms of E(M) for k = 2, 3 and fixed probabilities.

Table (9.1)

The effect of δ_0 on the performance characteristics of the schemes using the stopping rule MS_1 (δ_0) when N = 50, generated p_1 and p_2 with uniform prior (1, 1; 3).

Sampling rule	δ_0	Performance characteristics			
		P(CS)	E(M)	P(M < N)	P(CO)
MR_1	0.3	.8211	12.1680	.8498	.5291
	0.4	.8796	24.1680	.6176	.6337
	0.5	.8792	28.7753	.5014	.6332
	0.6	.9028	39.0071	.2877	.7219
	0.7	.8978	44.4612	.1551	.7473
MR_2 (5)	0.3	.8700	20.7635	.7233	.6432
	0.4	.9020	32.8290	.4680	.7043
	0.5	.8978	36.6505	.3597	.7112
	0.6	.8980	39.5185	.2723	.7221
	0.7	.9034	46.1745	.1230	.7698

NB: It is not worth looking at very small (near 0.0) or very large (near 1.0) values of δ_0 since the first case gives very small E(M) and consequently small P(CS) and the second case gives very large E(M) with roughly same P(CS).

Table (9.2)

The effect of f^* on the performance characteristics of the schemes using the stopping rule MS_2 (f^*) when N = 50, generated p_1 and p_2 with uniform prior (1, 1; 3).

Sampling rule	f^*	Performance characteristics			
		P(CS)	E(M)	P(M < N)	P(CO)
MR_1	1	.6106	1.0000	1.0000	.3043
	3	.8327	8.3768	.9987	.5578
	5	.8830	17.1505	.9396	.6493
	10	.8998	31.4231	.7043	.7291
	15	.9017	39.0270	.5339	.7575
	25	.9044	46.6227	.2744	.7774
MR_2 (5)	1	.7628	5.8360	1.0000	.5289
	3	.8492	12.3360	.9843	.6153
	5	.8914	20.5535	.8926	.6743
	10	.8994	33.3915	.6620	.7300
	15	.9012	40.5360	.4798	.7653
	25	.8977	47.1120	.2328	.7782

Table (9.3)

Performance characteristics of the schemes when k = 2 (trinomial case), generated p_1 and p_2, N = 10, 50, 100 with a uniform prior D(1, 1; 3).

Sampling rule	Stopping rule	N	Performance characteristics			
			P(CS)	E(M)	P(M < N)	P(CO)
MR_1	MS_1 (.3)	10	.7664	5.0779	.7346	.4601
		50	.8211	12.1680	.8498	.5291
		100	.8267	19.4718	.8608	.5376
	MS_2 (3)	10	.7864	6.4184	.7208	.5022
		50	.8327	8.3768	.9987	.5578
		100	.8321	8.3914	1.0000	.5567
	MS_3	10	.7913	7.3751	.9277	.5518
		50	.8954	38.4292	.9799	.7646
		100	.9291	77.2382	.9906	.8375
	MS_4	10	.7227	3.1529	1.0000	.4164
		50	.8875	21.7831	1.0000	.6736
		100	.9281	47.1669	1.0000	.7704
MR_2 (5)	MS_1 (.3)	10	.7940	7.8530	.4294	.5551
		50	.8700	20.7635	.7233	.6432
		100	.8979	33.3730	.7546	.6720
	MS_2 (3)	10	.7940	7.8530	.4294	.5551
		30	.8492	12.3360	.9843	.6153
		50	.8494	12.5020	.9998	.6191

Table (9.3) continued

Performance characteristics of the schemes when k = 2 (trinomial case), generated p_1 and p_2, N = 10, 50, 100 with a uniform prior D(1, 1; 3).

Sampling rule	Stopping rule	N	Performance characteristics			
			P(CS)	E(M)	P(M < N)	P(CO)
	MS_3	10	.7929	10.0000	.0000	.5805
		50	.9036	49.9110	.0178	.7922
		100	.9316	99.7710	.0298	.8502
	MS_4	10	.7502	5.6860	.8628	.5117
		50	.8917	24.6480	.9685	.6908
		100	.9289	50.5320	.9844	.7752

NB: For MR_3,

N	P(CS)	P(CO)
10	.7929	.5805
50	.9056	.7926
100	.9330	.8452

Table (9.4)

Performance characteristics of the group sequential schemes for k = 2 (trinomial case) using $MR_2(g)$ sampling method in conjunction with various stopping rules and various group sizes when N = 100, generated p_1 and p_2 with uniform priors (1, 1; 3).

Stopping rule	g	Performance characteristics			
		P(CS)	E(M)	P(M < N)	P(CO)
MS_1 (.3)	1	.8267	19.4718	.8608	.5376
	5	.8979	33.3730	.7546	.6720
	10	.9175	41.4370	.6993	.7347
	25	.9290	61.1075	.5626	.7951
	50	.9305	75.2700	.4946	.8314
MS_2 (3)	1	.8321	8.3914	1.0000	.5567
	5	.8494	12.5020	.9998	.6191
	10	.8792	17.5480	.9980	.6818
	25	.9006	31.1375	.9811	.7594
	50	.9182	54.7050	.9059	.8110
MS_3	1	.9291	77.2382	.9906	.8375
	5	.9316	99.7710	.0298	.8502
	10	.9330	100.0000	.0000	.8452
	25	.9330	100.0000	.0000	.8452
	50	.9330	100.0000	.0000	.8452

Table (9.4) continued

Performance characteristics of the group sequential schemes for k = 2 (trinomial case) using $MR_2(g)$ sampling method in conjunction with various stopping rules and various group sizes when N = 100, generated p_1 and p_2 with uniform priors (1, 1; 3).

Stopping rule	g	Performance characteristics			
		P(CS)	E(M)	P(M < N)	P(CO)
MS_4	1	.9281	47.1669	1.0000	.7704
	5	.9289	50.5320	.9844	.7752
	10	.9297	53.4910	.9376	.7848
	25	.9314	61.4400	.8006	.8035
	50	.9299	73.5950	.5281	.8277

Table (9.5)

Performance characteristics of the sampling schemes for k = 2 (trinomial case) when N = 50, fixed p_1 and p_2 with a uniform prior D(1, 1; 3).

$(p_{[1]}, p_{[2]}, p_{[3]})$	Sampling rule	Stopping rule	Performance characteristics			
			P(CS)	E(M)	P(M < N)	P(CO)
(.3, .3, .4)	MR_1	MS_1 (.3)	.5938	24.9075	.5572	.4054
		MS_2 (3)	.5810	14.4521	.9810	.3924
		MS_3	.6878	45.6867	.9296	.3825
		MS_4	.6713	33.7841	1.0000	.4004
	MR_2 (5)	MS_1 (.3)	.6538	38.3850	.2846	.3629
		MS_2 (3)	.6165	21.4690	.9192	.3695
		MS_3	.6892	50.0000	.0000	.3585
		MS_4	.6809	37.3670	.9406	.3941
(.15, .4, .45)	MR_1	MS_1 (.3)	.5946	20.0185	.6713	.5577
		MS_2 (3)	.5803	11.5050	.9961	.5218
		MS_3	.6526	44.9888	.9418	.6418
		MS_4	.6446	31.7108	1.0000	.6186
	MR_2 (5)	MS_1 (.3)	.6357	33.3440	.4102	.5683
		MS_2 (3)	.6153	17.3535	.9618	.5498
		MS_3	.6513	50.0000	.0000	.6161
		MS_4	.6464	35.3755	.9141	.6153
(.2, .2, .6)	MR_1	MS_1 (.3)	.9081	11.0026	.9280	.7021
		MS_2 (3)	.9474	8.2223	1.0000	.6688
		MS_3	.9990	37.0860	.9994	.5596
		MS_4	.9971	18.8808	1.0000	.6479

Table (9.5) continued

Performance characteristics of the sampling schemes for $k = 2$ (trinomial case) when $N = 50$, fixed p_1 and p_2 with a uniform prior $D(1, 1; 3)$.

$(p_{[1]}, p_{[2]}, p_{[3]})$	Sampling rule	Stopping rule	Performance characteristics			
			P(CS)	E(M)	P(M < N)	P(CO)
	MR_2 (5)	MS_1 (.3)	.9838	18.6470	.8519	.6098
		MS_2 (3)	.9736	11.3775	.9986	.5970
		MS_3	.9994	50.0000	.0000	.5484
		MS_4	.9976	21.3620	.9989	.6268

NB: For MR_3,

$(p_{[1]}, p_{[2]}, p_{[3]})$	P(CS)	P(CO)
(.3, .3, .4)	.6892	.3584
(.15, .4, .45)	.6513	.6161
(.2, .2, .6)	.9994	.5484

Table (9.6)

Performance characteristics of the schemes when $k = 3$, generated p_1, p_2 and p_3, $N = 10, 50, 100$ with a uniform prior $D(1, 1, 1; 4)$.

Sampling rule	Stopping rule	N	Performance characteristics			
			P(CS)	E(M)	P(M < N)	P(CO)
MR_1	MS_1 (.3)	10	.6991	6.0389	.5999	.2008
		50	.7811	18.4014	.7290	.3018
		100	.8001	31.7802	.7393	.3233
	MS_2 (3)	10	.7191	7.2481	.5945	.2345
		50	.7921	10.6552	.9939	.2819
		100	.7933	10.7080	1.0000	.2822
	MS_3	10	.7196	7.8756	.8734	.2557
		50	.8635	40.5598	.9723	.5409
		100	.9018	81.4461	.9858	.6411
	MS_4	10	.5180	1.0000	1.0000	.0843
		50	.8505	23.8133	1.0000	.4228
		100	.9017	52.8534	1.0000	.5624
MR_2 (5)	MS_1 (.3)	10	.7218	8.5455	.2909	.2582
		50	.8487	29.6820	.5331	.4166
		100	.8851	51.2410	.5731	.4750
	MS_2 (3)	10	.7218	8.5455	.2909	.2582
		50	.8196	15.2645	.9655	.3445
		100	.8204	15.7325	.9984	.3490

Table (9.6) continued

Performance characteristics of the schemes when $k = 3$, generated p_1, p_2 and p_3, $N = 10, 50, 100$ with a uniform prior $D(1, 1, 1; 4)$.

Sampling rule	Stopping rule	N	Performance characteristics			
			P(CS)	E(M)	P(M < N)	P(CO)
	MS_3	10	.7296	10.0000	.0000	.2905
		50	.8643	49.9855	.0029	.5749
		100	.8992	99.9730	.0047	.6640
	MS_4	10	.6808	6.0320	.7936	.2178
		50	.8556	27.0760	.9701	.4427
		100	.9020	56.6625	.9818	.5689

NB: For MR_3,

N	P(CS)	P(CO)
10	.7296	.2905
50	.8667	.5642
100	.8971	.6758

Table (9.7)

Performance characteristics of the schemes for $k = 3$, generated p_1, p_2 and p_3, $N = 10, 50, 100$ with Dirichlet prior $D(1, 2, 3; 10)$.

Sampling rule	Stopping rule	N	Performance characteristics P(CS)	E(M)	P(M < N)	P(CO)
MR_1	MS_1 (.3)	10	.7045	8.9408	.2124	.3352
		50	.8373	35.6065	.3870	.5042
		100	.8722	65.0930	.4267	.5639
	MS_2 (3)	10	.7034	6.5102	.6393	.3251
		50	.7599	9.5943	.9929	.3768
		100	.7588	9.7142	.9998	.3772
	MS_3	10	.7126	7.7774	.8815	.3523
		50	.8332	41.8371	.9665	.5514
		100	.8794	84.3933	.9829	.6401
	MS_4	10	.6472	2.6972	1.0000	.2856
		50	.8176	26.4127	1.0000	.4739
		100	.8709	58.3276	1.0000	.5883
MR_2 (5)	MS_1 (.3)	10	.7100	9.5955	.0809	.3549
		50	.8334	39.1270	.3163	.5205
		100	.8741	70.3725	.3803	.5798
	MS_2 (3)	10	.7111	8.3275	.3345	.3436
		50	.7834	14.9040	.9673	.4097
		100	.7865	15.4070	.9987	.4107

Table (9.7) continued

Performance characteristics of the schemes for $k = 3$, generated p_1, p_2 and p_3, $N = 10, 50, 100$ with Dirichlet prior $D(1, 2, 3; 10)$.

Sampling rule	Stopping rule	N	Performance characteristics			
			P(CS)	E(M)	P(M < N)	P(CO)
	MS_3	10	.7037	10.0000	.0000	.3542
		50	.8319	49.9975	.0005	.5628
		100	.8775	99.9980	.0003	.6618
	MS_4	10	.6850	5.8695	.8261	.3267
		50	.8303	29.4585	.9703	.4970
		100	.8725	61.5880	.9827	.5941

NB: For MR_3,

N	P(CS)	P(CO)
10	.7037	.3542
50	.8292	.5660
100	.8794	.6675

Table (9.8)

The exact (Kulkarni 1981) and approximate results of E(M) for $k = 2, 3$ with fixed probabilities under (MR_1, MS_3).

k	$(p_{[1]}, \ldots, p_{[k+1]})$	N	Exact	Approximate
2	(.33, .33, .33)	10	8.52	8.5553
		20	17.91	17.9333
	(.25, .25, .5)	10	8.16	8.1651
		20	16.69	16.6720
3	(.25, .25, .25, .25)	10	8.82	8.8110
		20	18.32	18.3085
	(.2, .2, .2, .4)	10	8.56	8.5544
		20	17.48	17.4515

9.5 Concluding remarks

Among all stopping rules; MS_2 (3) produces substantial reductions in the sample sizes while it achieves P(CS) close to others. The results of group sequential schemes show that for small values of the group size g a great reduction in E(M) can be gained with little decrease in P(CS). Consequently, on balancing these two criteria it would be preferable to use moderate values for g such as 4, 5, 10 out of 100.

CHAPTER 10

SUPPLEMENTARY INVESTIGATION AND FUTURE WORK

There are several directions in which further investigation is possible; some of which are listed below.

A - Binomial problem:

(i) The suboptimal schemes which are based on $\delta = \left|\frac{r_1}{n_1} - \frac{r_2}{n_2}\right|$, described in chapter 5, can be extended to the problem of selecting the best of $k \geq 3$ Binomial populations. In addition to the performance characteristics which have already been discussed in previous chapters using Monte Carlo simulation, some other characteristics such as the probability of selecting the inferior populations, the expected number of observations on each population $(E(N_{(i)}), i = 1, \ldots, k)$ can also be calculated under the case $k \geq 3$.

(ii) When $k \geq 3$ the above procedures can be developed further by eliminating the inferior populations during the experimentation successively. One way of performing this can be described as follows:

(a) Suppose that the observations are taken sequentially from the populations, singly or in groups with n observation on each population.

(b) Calculate the posterior estimate of p_i, denoted by $\hat{p}_i$,

$\frac{r_i}{n_i}$ $(i = 1, \ldots, k)$.

(c) Eliminate population π_i from further consideration if

$$\hat{p}_{[k]} - \hat{p}_i > \delta', \qquad (10.1)$$

where δ' $0 < \delta' < 1$ is preassigned.

(d) If only one population remains we stop sampling and announce that this population is the best. Otherwise we continue sampling by taking another single observation (or group of observations) from the remaining populations and repeat the steps (b) and (c). Proceeding this way until only one population is left at which time the procedure is terminated with the declaration that this remaining population is the best. Ties are broken by randomization. If at any stage of the experimentation all the populations left are eliminated, then population π_j, whose posterior estimate $\hat{p}_j$ is the highest among these populations, is selected as the best population.

(iii) Bayesian sequential schemes for selecting the better of two Binomial populations with stopping rules based on the posterior probability that $p_1 > p_2$ and different types of stopping boundaries can be developed. In the following we describe these schemes.

Let N be the maximum total number of observations,

M be the actual total number of observations taken from the populations during the procedure,

δ_0 be a preassigned value such that $0 < \delta_0 < 0.5$.

The rules of the schemes are:

Sampling rule: The observations are taken sequentially one at a time or in groups with $\hat{p}_1$ and $\hat{p}_2$ calculated by r_1/n_1 and r_2/n_2 respectively. Then the sampling decisions are as follows:

If $\hat{p}_1 > \hat{p}_2$ sample from π_1

$\hat{p}_1 < \hat{p}_2$ sample from π_2

$\hat{p}_1 = \hat{p}_2$ sample at random.

Stopping rule: At the point (r_1, n_1, r_2, n_2), let $P(r_1, n_1, r_2, n_2)$ be the posterior probability that $p_1 > p_2$, given in section 8.1. Further, let $a(M)$ and $b(M)$ be the upper and lower stopping boundary respectively. Then,

stop sampling and proceed to the terminal decision rule if one of the following boundary conditions

$$P(r_1, n_1, r_2, n_2) > a(M), \text{ or} \tag{10.2}$$

$$P(r_1, n_1, r_2, n_2) < b(M) \tag{10.3}$$

is satisfied.

Terminal decision rule:

Take decision D_2: $p_1 > p_2$ if (10.2) is satisfied,

and D_1: $p_1 \leq p_2$ if (10.3) is satisfied.

Three types of stopping boundaries are proposed. These are:

1. Parallel stopping boundaries PB: Where

$a(M) = 0.5 + \delta_0$,

$b(M) = 0.5 - \delta_0$.

2. Intersected boundaries IB_1: Where

$$a(M) = (0.5 + \delta_0) - \frac{\delta_0 M}{N},$$

$$b(M) = (0.5 - \delta_0) + \frac{\delta_0 M}{N},$$

3. Intersected boundaries IB_2:

It is a special case of IB_1 where $\delta_0 = 0.5$. The boundaries are

$$a(M) = 1 - \frac{M}{2N},$$

$$b(M) = \frac{M}{2N}.$$

Graphically, Figs. (10.1 - 10.3) show the shape of these boundaries.

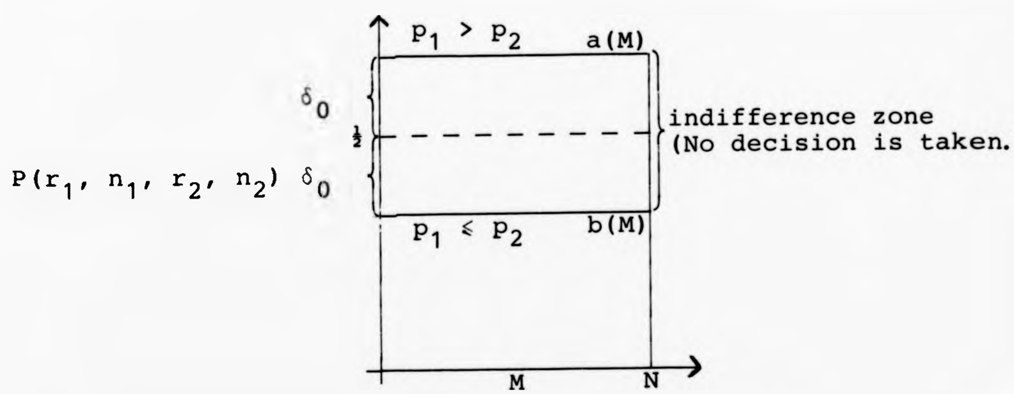

Fig. (10.1) Parallel boundaries (PB).

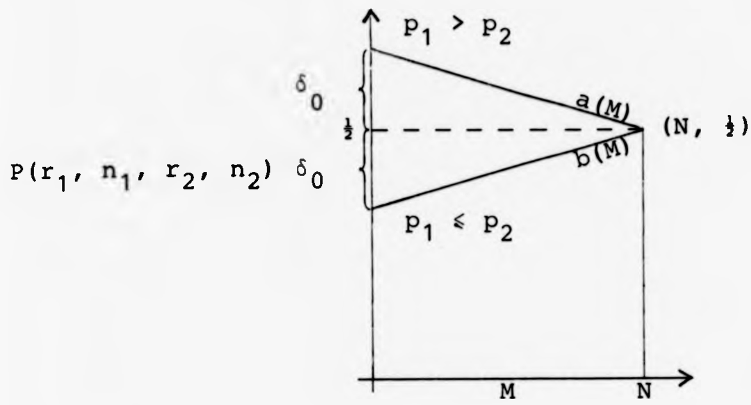

Fig. (10.2) Intersected boundaries (IB_1).

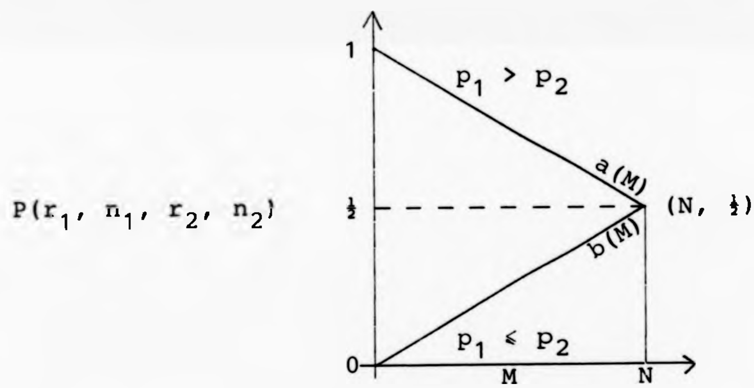

Fig. 10.3) Intersected boundaries (IB_2).

Brief numerical investigation of the schemes, using Monte Carlo simulation, has been carried out. Some performance characteristics such as P(CS), E(M), $E(N_{(1)})$, P(NOD) (the probability of no decision is taken) and P(E) (the probability of error) have been calculated under uniform priors; the results are presented in Tables (10.1 - 10.2).

From Table (10.1) it is clear that P(CS) decreases (increases) as δ_0 increases under the stopping boundary conditions PB (IB_1). On the other hand, P(NOD) increases rapidly (slowly) as δ_0 increases under the stopping boundary conditions PB (IB_1). The measures E(M) and $E(N_{(1)})$ increase as δ_0 increases but the rate of increase under PB is greater than IB_1. The measure P(E) decreases rapidly (slowly) as δ_0 increases under PB (IB_1).

Table (10.2) presents some results on the performance of the schemes for various values of N with $\delta_0 = 0.2$. The results show that IB_1 is slightly better than PB in terms of E(M) and $E(N_{(1)})$. As expected, P(NOD) is zero under IB_1 and greater than zero under PB. If we take into account P(NOD) then it is not sensible to measure the performance of the schemes by P(CS); however, P(E) will be the reasonable alternative according to which PB is better than IB_1. It is noted from the same table that the performance measures are not very sensitive to the sample size N under both types of stopping boundaries, PB and IB_1.

Table (10.1)

The effect of δ_0 on the performance characteristics of the schemes under two types of stopping boundary conditions PB and IB_1 with N = 20 and uniform priors.

Boundary conditions	δ_0	Performance characteristics					
		P(CS)	P(E)	P(NOD)	E(M)	$E(N_{(1)})$	E(R)
PB	0.2	.7624	.2237	.0139	3.6237	2.0515	1.9892
	0.3	.7010	.1493	.1497	7.2874	4.5449	3.7196
	0.4	.4622	.0434	.4944	13.9793	9.6754	6.2512
	0.5	.0000	.0000	1.0000	20.0000	15.1313	7.6534
IB_1	0.2	.7638	.2362	.0000	3.2418	1.8241	1.7519
	0.3	.7922	.2078	.0000	4.7596	2.8202	2.4469
	0.4	.7983	.2015	.0002	6.6837	4.1367	3.3281
	0.5	.8076	.1782	.0142	8.9653	5.8525	4.2295

NB: At $\delta_0 = 0.5$, IB_1 is equivalent to IB_2.

Table (10.2)

Performance characteristics of the schemes under two types of stopping boundary conditions PB and IB_1 with $\delta_0 = 0.2$, N = 10(10)100 and uniform priors.

Boundary conditions	N	Performance characteristics					
		P(CS)	P(E)	P(NOD)	E(M)	$E(N_{(1)})$	E(R)
PB	10	.7351	.2143	.0506	3.3875	1.9223	1.7977
	20	.7624	.2237	.0139	3.6237	2.0515	1.9892
	30	.7669	.2263	.0068	3.7239	2.1132	2.0701
	40	.7686	.2284	.0030	3.7679	2.1387	2.1030
	50	.7702	.2283	.0015	3.7694	2.1414	2.1037
	60	.7706	.2282	.0012	3.7844	2.1531	2.1134
	70	.7709	.2279	.0012	3.8112	2.1681	2.1373
	80	.7709	.2280	.0011	3.8238	2.1761	2.1476
	90	.7718	.2274	.0008	3.8356	2.1801	2.1556
	100	.7711	.2282	.0007	3.8333	2.1844	2.1529
IB_1	10	.7594	.2406	.0000	3.0568	1.7246	1.5921
	20	.7638	.2362	.0000	3.2418	1.8241	1.7519
	30	.7641	.2359	.0000	3.4049	1.9075	1.8688
	40	.7641	.2359	.0000	3.4334	1.9215	1.8955
	50	.7670	.2330	.0000	3.4466	1.9402	1.8967
	60	.7669	.2331	.0000	3.4505	1.9403	1.8990
	70	.7671	.2329	.0000	3.4585	1.9452	1.9048
	80	.7664	.2336	.0000	3.4621	1.9464	1.9082
	90	.7655	.2345	.0000	3.4641	1.9469	1.9095
	100	.7658	.2342	.0000	3.4605	1.9461	1.9053

B - Multinomial problem:

An optimal sequential scheme to select the best cell in the multinomial distribution can be found using a Bayesian decision theoretic approach. As given in chapter 9, consider a multinomial distribution with $(k + 1)$ cells with unknown probability of an observation in the i^{th} cell p_i $(i = 1, \ldots, k)$, where $p_{k+1} = 1 - \sum_{i=1}^{k} p_i$. Further, suppose that

$$p_{[1]} < p_{[2]} < \cdots < p_{[k+1]} \tag{10.4}$$

are the ordered values of $p_1, \ldots, p_{k+1}$.

Consider the following terminal decision rule:

$$d_i': p_i \text{ is the largest,} \qquad (i = 1, \ldots, k + 1). \tag{10.5}$$

Define the loss function by

$$\begin{aligned} L(p_{[k+1]} - p_i) &= K_m(p_{[k+1]} - p_i) && \text{if } p_{[k+1]} > p_i, \\ &= 0 && \text{if } p_{[k+1]} = p_i, \end{aligned} \tag{10.6}$$

where K_m is the loss constant.

Let $p_1, p_2, \ldots, p_{k+1}$ are assigned the Dirichlet distribution whose density function is given by (9.1.1). After m observations with n_i in the cell i $(i = 1, \ldots, k)$ the posterior distribution of p_i is Dirichlet with parameters $(n_i' + n_i = n_i'', \ i = 1, \ldots, k)$ and $(m' + m = m'')$ with mean $(\hat{p}_i = n_i'' / m'')$. The maximum sample size N can be taken through the procedure.

The stopping risk of taking decision d_i', denoted by

$S_i(n_1, n_2, \ldots, n_{k+1}; m)$, can be found as follows:

$$S_i(n_1, n_2, \ldots, n_{k+1}; m) = E[L(p_{[k+1]} - p_i)] \tag{10.7}$$

$$= K_m[E(p_{[k+1]}) - \hat{p}_i]. \tag{10.8}$$

However, it is difficult to find the posterior expected value of $p_{[k+1]}$.

At the point $(n_1, n_2, \ldots, n_{k+1}; m)$, let

c' be the cost of sampling one observation,

$B(n_1, n_2, \ldots, n_{k+1}; m)$ be the risk of taking one further observation and proceeding optimality thereafter (continuation risk),

$D(n_1, n_2, \ldots, n_{k+1}; m)$ be the optimal risk.

At each point in the $(k + 1)$ integer space, there are $(k + 1)$ possible transitions, $(\underset{\sim}{n} + \underset{\sim}{e}_i; m + 1)$ with probability $\hat{p}_i$ where $\underset{\sim}{e}_i = (0, \ldots, \underset{i^{th}\text{ element}}{\underset{\downarrow}{1}}, 0, 0)$ and $i = 1, \ldots, k + 1$.

Then the dynamic programming equations for the procedure are:

$$B(n_1, n_2, \ldots, n_{k+1}; m) = c' + \sum_{i=1}^{k+1} \hat{p}_i \, D(\underset{\sim}{n} + \underset{\sim}{e}_i; m + 1), \tag{10.9}$$

and

$$D(n_1, n_2, \ldots, n_{k+1}; m) = \min[S_1(n_1, n_2, \ldots, n_{k+1}; m), \ldots,$$

$$S_{k+1}(n_1, n_2, \ldots, n_{k+1}; m),$$

$$B(n_1, n_2, \ldots, n_{k+1}; m)]. \qquad (10.10)$$

REFERENCES

Alam, K. (1971). On selecting the most probable category. Technometrics, Vol. 13, No. 4, pp. 843-850.

Alam, K. and Thompson, J.R. (1972). On selecting the least probable multinomial event. Ann. Math. Statist., Vol. 43, No. 6, pp. 1981-1990.

Anscombe, F.J. (1963). Sequential medical trials. J. Amer. Statist. Assoc., Vol. 58, pp. 365-383.

Arkles, L. and Srinivasan, R. (1979). On the sequential selection of the better of two Binomial populations. Sankhyā, Vol. 41, Ser. B, pts. 1 and 2, pp. 15-30.

Armitage, P. (1975). Sequential medical trials. Oxford, Blackwell.

Atkinson, A.C. and Pearce, M.C. (1976). The computer generation of Beta, Gamma and Normal random variables. J. Roy. Statist. Soc. Ser. A , Pt. 4, pp. 139.

Bechhofer, R.E. (1954). A single-sample multiple decision procedure for ranking means of Normal populations with known variances. Ann. Math. Statist., 25, pp. 16-39.

Bechhofer, R.E., Elmaghraby, S. and Morse, N. (1959). A single-sample multiple-decision procedure for selecting the multinomial event which has the highest probability. Ann. Math. Statist., 30, pp. 102-119.

Bechhofer, R.E. and Frisardi, T. (1982). A Monte Carlo study of the performance of a closed adaptive sequential procedures for selecting the best Bernoulli population. T.R. No. 560, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York.

Bechhofer, R.E., Kiefer, J. and Sobel, M. (1968). Sequential identification and ranking procedures. The University of Chicago Press, Chicago.

Bechhofer, R.E. and Kulkarni, R.V. (1981). Closed adaptive sequential procedures for selecting the best of $k \geq 2$ Bernoulli populations. T.R. No. 510, School of Operations Research and Industrial Engineering, College of Engineering, Cornell University, Ithaca, New York.

Bechhofer, R.E. and Kulkarni, R.V. (1982). On the performance characteristics of a closed adaptive sequential procedure for selecting the best Bernoulli population. School of Operation Research and Industrial Engineering, Cornell University, Ithaca, New York.

Bellman, R.E. (1957). Dynamic programming. Rand Corporation Research Studies, Princeton, N.J.

Berger, R.L. (1980). Minimax subset selection for the Multinomial distribution. Journal of Statistical Planning and Inference 4, pp. 391–402.

Berry, D.A. (1972). A Bernoulli two-armed bandit. Ann. Math. Statist., Vol. 43, No. 3, pp. 871-897.

Berry, D.A. and Sobel, M. (1973). An improved procedure for selecting the better of two Bernoulli populations. J. Amer. Statist. Assoc., Vol. 68, No. 344, pp. 979-984.

Bland, R.P. and Bratcher, T.L. (1968). A Bayesian approach to the problem of ranking Binomial probabilities. SIAMJ. Appl. Math., Vol. 16, No. 4, pp. 843-850.

Bratcher, T.L. and Bland, R.P. (1975). On comparing Binomial probabilities from a Bayesian viewpoint. Comm. Statist., 4(10), pp. 975-985.

Büringer, H., Martin, H. and Schriever, K.H. (1980). Nonparametric sequential selection procedures. Birkhauser, Boston, Massachusetts.

Cacoullos, T. and Sobel, M. (1966). An inverse-sampling procedure for selecting the most probable event in a Multinomial distribution. In P.R. Krishnaiah (ed.), Multivariate Analysis. New York. Academic Press.

Canner, P.C. (1970). Selecting one of two treatments when responses are dichotomous. J. Amer. Statist. Assoc., Vol. 65, pp. 293-306.

Cheng, R.C.H. (1977). The generation of Gamma variables with non-integer shape parameter. Appl. Statist., Vol. 26, No. 1, pp. 71-75.

Colton, T. (1963). A model for selecting one of two medical treatments. J. Amer. Statist. Assoc., Vol. 58, pp. 388-400.

Dudewicz, E.J. (1976). Introduction to statistics and probability. Holt, Rinehart and Winston.

Dudewicz, D.J. and Koo, Joo Ok (1982). The complete categorized guide to statistical selection and ranking procedures. American Sciences Press INC., Columbus, Ohio, USA.

El-Sayyad, G.M. and Freeman, P.R. (1973). Bayesian sequential estimation of a Poisson process rate. Biometrika, 60, pp. 289-296.

Fabius, J. and Van Zwet, W.R. (1970). Some remarks on the two-armed bandit. Ann. Math. Statist., Vol. 41, No. 6, pp 1906-1916.

Freeman, P.R. (1970). Optimal Bayesian sequential estimation of the median effective dose. Biometrika, 57, 1, pp. 79-89.

Freeman, P.R. (1972). Sequential estimation of the size of a population. Biometrika, 59, 1, pp. 9-17.

Fushimi, M. (1973). An improved version of a Sobel-Weiss play-the-winner procedure for selecting the better of two Binomial populations. Biometrika, 60, 3, pp. 517-523.

Gibbons, J.D., Olkin, I. and Sobel, M. (1977). Selecting and ordering populations: A new statistical methodology. John Wiley and Sons, New York.

Gibbons, J.D., Olkin, I. and Sobel, M. (1979). An introduction to ranking and selection. The American Statistican, Vol. 33, No. 4, pp. 185-195.

Goel, P.K. and Rubin, H. (1977). On selection a subset containing the best population - A Bayesian approach. Ann. Statist., Vol. 5, No. 5, pp. 969-983.

Gupta, S.S. and Huang, D-Y (1976). On subset selection procedures for the entropy function associated with the Binomial populations. Sankhyā, Ser. A, Vol. 38, Pt. 2, pp. 153-173.

Gupta, S.S., Huyett, M.J. and Sobel, M. (1957). Selection and ranking with Binomial populations. Trans. Amer. Soc. Qual. Contr. pp. 635-644.

Gupta, S.S. and Nagel, K. (1967). On selection and ranking procedures and order statistics from the Multinomial distribution. Sankhyā, Ser. B, Vol. 29, Parts 1 and 2, pp 1-34.

Gupta, S.S. and Panchapakesan, S. (1979). Multiple decision procedures: Theory and methodology of selecting and ranking populations. John Wiley and Sons, N.Y.

Hoel, D.G. (1972). An inverse stopping rule for play-the-winner sampling. J. Amer. Statist. Assoc., Vol. 67, No. 337, pp. 148-151.

Hoel, D.G. and Sobel, M. (1972). Comparisons of sequential procedures for selecting the best Binomial populations. Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. IV (L. LeCam, J. Neyman and E.L. Scott, eds.), Univ. of California Press, Berkeley and Los Angeles, pp. 53-69.

Hoel, D.G., Sobel, M. and Weiss, G.H. (1972). A two-stage procedure for choosing the better of two Binomial populations. Biometrika, 59, 2, pp. 317-322.

Hoel, D.G., Sobel, M. and Weiss, G.H. (1975a). Comparison of sampling methods for choosing the best Binomial population with delayed observations. J. Statist. Comput. Simul., Vol. 3, pp. 299-313.

Hoel, D.G., Sobel, M. and Weiss, G.H. (1975b). A survey of adaptive sampling for clinical trials. Perspectives in Biometry (Ed. R.M. Elashoff), Academic, New York, pp. 29-61.

Hwang, F.K. (1982). A multistage selection scheme for the most probable event. American Journal of Mathematical and Management Sciences, Vol. 2, No. 1, pp. 3-12.

Jennison, C. (1983). Equal probability of correct selection for Bernoulli selection procedures. Commun. Statist.- Theor. Meth., 12(24), pp. 2887-2896.

Jennison, C. (1984). On the expected sample size for the Bechhofer-Kulkarni Bernoulli selection procedure. Sequential Analysis, 3(1), pp. 39-49.

Jones, P.W. (1974). A note on the Bayesian sequential estimation of a Binomial parameter. Biometrika, 61, pp. 642-644.

Jones, P.W. (1975). The two-armed bandit. Biometrika, 62, 2, pp. 523-524.

Jones, P.W. and Kandeel, H.A. (1984). Numerical investigation of the two-armed bandit. Mathematical learning models - Theory and Algorithm (Springer lecture notes in statistics No. 20) (Eds. U. Herkenrath, D. Kalin and W. Vogel) pp. 101-107.

Kiefer, J.E. and Weiss, G.H. (1971). A truncated test for choosing the better of two Binomial populations. J. Amer. Statist. Assoc., Vol. 66, No. 336, pp. 867-871.

Kiefer, J.E. and Weiss, G.H. (1974). Truncated version of a play-the-winner rule for choosing the better of two Binomial populations. J. Amer. Statist. Assoc., Vol. 69, No. 347, pp. 807-809.

Kulkarni, R.V. (1981). Closed adaptive sequential procedures for selecting the best of $k \geq 2$ Bernoulli populations. Ph.D. dissertation, Cornell University, Ithaca, New York.

Kulkarni, R.V. and Jennison, C. (1983). Optimal properties of the Bechhofer-Kulkarni Bernoulli selection procedure. T.R. No. 600, School of Operations Research and Industrial Engineering, College of Engineering, Cornell University, Ithaca, New York. To appear in Ann. Statist.

Kulkarni, R.V. and Kulkarni, V.G. (1986). Optimal Bayes procedures for selecting the better of two Bernoulli populations. To be published in Journal of Statistical Planning and Inference.

Lindley, D.V. (1961). Dynamic programming and decision theory. Appl. Statist., Vol. 10, No. 2, pp. 39-51.

Lindley, D.V. and Barnett, B.N. (1965). Sequential sampling: two decision problems with linear losses for Binomial and Normal random variables. Biometrika, 52, 3 and 4, pp. 507-532.

Meeter, D.A. (1975). A two-armed bandit with terminal decision (Bayes rules). J.N. Srivastava, ed., A survey of statistical design and linear models. North-Holland Publishing Company.

NAGFLIB: 1454/0: MK6: May 77, Numerical Algorithms Group Ltd., Oxford, U.K.

Nebenzahl, E. and Sobel, M. (1972). Play-the-winner sampling for a fixed sample size Binomial selection problem. Biometrika, 59, 1, pp. 1-8.

Paulson, E. (1967). Sequential procedures for selecting the best of several Binomial populations. Ann. Math. Statist., Vol. 38, pp. 117-123.

Pocock, S.T. (1977). Group sequential methods in the design and analysis of clinical trials. Biometrika, 64, 2, pp. 191-199.

Poloniecki, J.D. (1979). The two-armed bandit and the controlled clinical trial. The Statistician, Vol. 27, No. 2, pp. 97-102.

Raiffa, H. and Schlaifer, R. (1968). Applied statistical decision theory. Massachusetts Institute of Technology Press.

Robbins, H. (1956). A sequential design problem with a finite memory. Proc. Nat. Acad. Sci. U.S.A. 42, pp. 920-923.

Ramey, T.J. and Alam, K. (1979). A sequential procedure for selecting the most probable multinomial event. Biometrika, 66, 1, pp. 171-173.

Ramey, T.J. and Alam, K. (1980). A Bayes sequential procedure for selecting the most probable multinomial event. Commun. Statist. - Theor. Meth., A9(3), pp. 265-276.

Simon, R., Weiss, G.H. and Hoel, D.G. (1975). Sequential analysis of Binomial clinical trials. Biometrika, 62, 1, pp. 195-200.

Simpson, M.G. (1961). An introduction to dynamic programming. Appl. Statist., Vol. 10, No. 1, pp. 32-38.

Sobel, M. and Huyett, M.J. (1957). Selecting the best one of several Binomial populations. Bell System Tech. J. 36, pp. 537-576.

Sobel, M. and Weiss, G.H. (1970). Play-the-winner sampling for selecting the better of two Binomial populations. Biometrika, 57, 2, pp. 357-365.

Sobel, M. and Weiss, G.H. (1972). Play-the-winner rule and inverse sampling for selecting the best of $k \geq 3$ Binomial populations. Ann. Math. Statist., Vol. 43, No. 6, pp. 1808-1826.

Tamhane, A.C. (1985). Some sequential procedures for selecting the better Bernoulli treatment by using a matched sample design. J. Amer. Statist., Assoc., Vol. 80, No. 390, pp. 455-460.

Taylor, R.J. and David, H.A. (1962). A multistage procedure for the selection of the best of several populations. J. Amer. Statist. Assoc., 57, pp. 785-795.

Wahrenberger, D.L., Antle, C.E. and Klimko, L.A. (1977). Bayesian rules for the two-armed bandit problem. Biometrika, 64, 1, pp. 172-174.

Wetherill, G.B. (1961). Bayesian sequential analysis. Biometrika, 48, 3 and 4, pp. 281-292.

Wilks, S.S. (1962). Mathematical statistics. John Wiley and Sons, New York.

Winkler, R.L. (1972). Introduction to Bayesian inference and decision. Holt, Rinehart and Winston, Inc.

Zelen, M.C. (1969). Play-the-winner rule and the controlled clinical trial. J. Amer. Statist. Assoc., 64, pp. 131-146.

APPENDICES

Appendix (3.1)

Listing of Program OPT1

This program has been developed to compute the overall risk, i.e. the risk at the origin, for the scheme OPT_1 under linear loss function.

The input data are:

(1) loss constants and sampling costs,

(2) prior information on p_1 and p_2.

The main variable names used in the program (with identifiers used in the text in brackets) are:

K1, K2	loss constants (K_1, K_2)
C1, C2	sampling costs (C_1, C_2)
RD1, ND1	prior number of successes, observations on population 1 (a_1, b_1)
RD2, ND2	prior number of successes, observations on population 2 (a_2, b_2)
R1, N1	actual number of successes, observations on population 2 (c_1, d_1)
R2, N2	actual number of successes, observations on population 2 (c_2, d_2)
S1, S2	stopping risks of taking decision D_1 and D_2 (S_1, S_2)
B1, B2	continuation risks with population 1 and 2 (B_1, B_2)
B	min(B_1, B_2)(B)
NW1 = 1, 2, 3, 4	type of risk ($S_1 \equiv 1$, $S_2 \equiv 2$, $B_1 \equiv 3$, $B_2 \equiv 4$)
S	array of optimal risks ($D(r_1, n_1, r_2, n_2)$)
N	maximum sample size (N).

```
      PROGRAM OPT1(INPUT,OUTPUT,TAPE2=INPUT,TAPE6=OUTPUT)
      REAL  K1,K2,IV(0:150)
      INTEGER  RD1,RD2,R1,R2,R4,R5,R6
      ROWWISE S(128,11,11)
      COMMON S
       IV(0)=1.0
       DO 133 J=1,150
       IV(J)=IV(J-1)*J
 133   CONTINUE
      READ(2,*,END=10000)K1,K2,C1,C2
10000 READ(2,*,END=10001)MM,ND1,ND2,RD1,RD2
      WRITE(6,99)
 99   FORMAT(5X,'RESULTS-OPT1')
      WRITE(6,100)
100   FORMAT(5X,'------------')
10001 WRITE(6,30)K1,K2,C1,C2
      WRITE(6,20)MM,ND1,ND2,RD1,RD2
C     -----------------------------------------------------------
C     -----------------------------------------------------------
      WRITE(6,61)
   61 FORMAT(5X,' MINIMUM RISK ',5X,'TYPE OF DECISIONS',5X,'N')
      WRITE(6,65)
   65 FORMAT(5X,'--------------',5X,'-----------------',5X,'-')
      DO 1 II=1,MM
      N=10*II
      NM=N+1
C-------------------------------------------------------------
C--    THE DO LOOP 11 COMPUTES THE STOPPING RISKS WHERE N=N1+N2.
C-------------------------------------------------------------
      DO 11 IM=1,NM
      N2=IM-1
      N1=N-N2
      NM1=N1+1
      NM2=N2+1
      N11=ND1+N1
      N22=ND2+N2
      LU1=N11-1
      LU2=N22-1
      Q=IV(LU1)*IV(LU2)
      Q1=K1*Q
      Q2=K2*Q
      R6=N22-1
      LR6=N11-1
      L8=N11+N22-1
      DO 11 I=1,NM1
      R1=I-1
      R4=L8-RD1-R1-1
      L6=N11-RD1-R1-1
      LR5=1-RD1-R1
      LL1=RD1+R1
      L7=LL1-1
      Q3=IV(L6)*IV(L7)*IV(L8)
      RQ3=1.0/Q3
      DO 11 L=1,NM2
      R2=L-1
      LL2=RD2+R2
      SUM=0.0
```

```
      R5=1-RD2-R2
      LL6=N22-RD2-R2-1
      LR4=L8-RD2-R2-1
      LL7=LL2-1
      Q4=IV(LL6)*IV(LL7)*IV(L8)
      RQ4=1.0/Q4
      DO 29 J=LL2,LU2
      L1=LL1+J
      L2=R4-J
      L3=R5+J
      L4=J+1
      L5=R6-J
      Q5=IV(L4)*IV(L5)
      RQ5=1.0/Q5
   29 SUM=SUM+IV(L1)*IV(L2)*L3*RQ5
      S1=Q1*SUM*RQ3
      NW1=1
      SUM=0.0
      DO 14 J=LL1,LU1
      L1=LL2+J
      L2=LR4-J
      L3=LR5+J
      L4=J+1
      L5=LR6-J
      Q5=IV(L4)*IV(L5)
      RQ5=1.0/Q5
   14 SUM=SUM+IV(L1)*IV(L2)*L3*RQ5
      S2=Q2*SUM*RQ4
      IF(S2.GE.S1)GOTO 11
      S1=S2
      NW1=2
   11 S(IM,I,L)=S1
C-----------------------------------------------------------------
C    ---THE DO LOOP 12 COMPUTES THE STOPPING AND  CONTINUATION
C            RISKS WHERE N1+N2=N-1  TO  N1+N2=0
C--------------------------------------------------------------
      DO 12 IN=1,N
      NN=N-IN
       NM=NN+1
      DO 13 IM=1,NM
      N2=IM-1
      N1=N-IN-N2
      N11=ND1+N1
      N22=ND2+N2
       RN11=1.0/(N11)
       RN22=1.0/(N22)
      NM1=N1+1
      NM2=N2+1
      LU1=N11-1
      LU2=N22-1
      Q=IV(LU1)*IV(LU2)
      Q1=K1*Q
      Q2=K2*Q
      R6=N22-1
      LR6=N11-1
      L8=N11+N22-1
      DO 17 I=1,NM1
```

```
      R1=I-1
      R4=L8-RD1-R1-1
      L6=N11-RD1-R1-1
      LL1=RD1+R1
      L7=LL1-1
      LR5=1-RD1-R1
      Q3=IV(L6)*IV(L7)*IV(L8)
      RQ3=1.0/Q3
      DO 17 L=1,NM2
      R2=L-1
      LR4=L8-RD2-R2-1
      LL6=N22-RD2-R2-1
      LL2=RD2+R2
      LL7=LL2-1
      R5=1-RD2-R2
       Q4=IV(LL6)*IV(LL7)*IV(L8)
       RQ4=1.0/Q4
      B1=C1+((RD1+R1)*S(IM,I+1,L)+
     1(N11-RD1-R1)*S(IM,I,L))*RN11
      B=B1
      NW2=3
      B2=C2+((RD2+R2)*S(IM+1,I,L+1)+(N22-RD2-R2)
     1*S(IM+1,I,L))*RN22
      IF(B1.LE.B2)GOTO 27
      B=B2
       NW2=4
27     SUM=0.
      DO 28 J=LL2,LU2
       L1=LL1+J
       L2=R4-J
        L3=R5+J
       L4=J+1
       L5=R6-J
       Q5=IV(L4)*IV(L5)
       RQ5=1.0/Q5
  28   SUM=SUM+IV(L1)*IV(L2)*L3*RQ5
       S1=Q1*SUM*RQ3
       NW1=1
       SUM=0.0
       DO 19 J=LL1,LU1
       L1=LL2+J
       L2=LR4-J
       L3=LR5+J
       L4=J+1
       L5=LR6-J
       Q5=IV(L4)*IV(L5)
       RQ5=1.0/Q5
 19      SUM=SUM+IV(L1)*IV(L2)*L3*RQ5
       S2=Q2*SUM*RQ4
       IF(S2.GE.S1)GOTO 22
       S1=S2
       NW1=2
 22    S(IM,I,L)=S1
      IF(S(IM,I,L).LE.B)GOTO 17
      NW1=NW2
      S(IM,I,L)=B
  17  CONTINUE
```

```
   13 CONTINUE
C-----------------------------------------------------------------
      IF(NN.NE.0)GOTO 12
      DO 23 IM=1,NM
      N2=IM-1
      N1=N-IN-N2
      NM1=N1+1
      NM2=N2+1
      DO 23 I=1,NM1
      R1=I-1
      DO 23 L=1,NM2
      R2=L-1
      WRITE(6,70)S(IM,I,L),NW1,N
   70 FORMAT(5X,E14.6,13X,I1,13X,I3)
   23 CONTINUE
   12 CONTINUE
    1 CONTINUE
   20 FORMAT(BZ,5I5)
   30 FORMAT(BZ,8E14.6)
      STOP
      END
```

Appendix (4.1)

Listing of Program OFSS

This program has been developed to compute the risk values of the scheme OFSS under linear loss function using equation (4.2.2). Further it can be used to find best values of N_1 ($N_2 = N - N_1$) which produce minimum risk for each particular N.

The input data are:

(1) loss constants and sampling costs,

(2) prior information on p_1 and p_2.

The main variable names used in the program (with identifiers used in the text in brackets) are:

K1, K2	loss constants (K_1, K_2)
C1, C2	sampling costs (C_1, C_2)
RD1, ND1	prior number of successes, observations on population 1 (a_1, b_1)
RD2, ND2	prior number of successes, observations on population 2 (a_2, b_2)
R1, N1	actual number of successes, observations on population 1 (c_1, N_1)
R2, N2	actual number of successes, observations on population 2 (c_2, N_2)
S1, S2	stopping risks of taking decision D_1 and D_2 (S_1, S_2)
P	the value of equation 4.2.3 at particular c_1, c_2 ($P(c_1, c_2)$)
NOPT1, NOPT2	best values of N_1, N_2
SUM2	overall risk of OFSS ($D_{FS}(N, N_1, N_2)$)
N	total sample size.

```
      PROGRAM OFSS(INPUT,OUTPUT,TAPE2=INPUT,TAPE6=OUTPUT)
      REAL  IV(0:150)
      INTEGER  RD1,RD2,R1,R2,R4,R5,R6,C1,C2
       IV(0)=1.0
       DO 133 J=1,150
       IV(J)=IV(J-1)*J
 133   CONTINUE
      READ(2,*,END=10000)K1,K2,C1,C2
10000 READ(2,*,END=10001)MM,ND1,ND2,RD1,RD2
      WRITE(6,99)
99    FORMAT(5X,'RESULTS-FIXED SAMPLE SIZE')
      WRITE(6,100)
100   FORMAT(5X,'-------------------------')
10001 WRITE(6,20)K1,K2,C1,C2
      WRITE(6,20)MM,ND1,ND2,RD1,RD2
C     -----------------------------------------------------------
C     -----------------------------------------------------------
      WRITE(6,61)
   61 FORMAT(5X,' EXPECTED  RISK ',10X,'N1',5X,'N2',9X,'N')
      WRITE(6,65)
   65 FORMAT(5X,'----------------',5X,'-----------------',5X,'-')
      ND11=ND1-1
      ND22=ND2-1
      RD11=RD1-1
      RD22=RD2-1
      NRD1=ND1-RD1-1
      NRD2=ND2-RD2-1
      DO 1 II=1,MM
      N=10*II
      NM=N+1
      SUM2=0.0
      SUM3=0.0
C-------------------------------------------------
C--   THE DO LOOP 2 COMPUTES THE RISKS FOR OFSS FOR
C--   ALL COMBINATIONS OF N1, N2 SUCH THAT N1+N2=N
C-------------------------------------------------
      DO 2 IM=1,NM
      N1=IM-1
      N2=N-N1
      NM1=N1+1
      NM2=N2+1
      N11=ND1+N1
      N22=ND2+N2
      LU1=N11-1
      LU2=N22-1
      Q=IV(LU1)*IV(LU2)
      Q1=K1*Q
      Q2=K2*Q
      R6=N22-1
      LR6=N11-1
      L8=N11+N22-1
      SUM1=0.0
C---------------------------------------------------
C--  THE DO LOOP 11 COMPUTES THE STOPPING RISKS S1 ,S2 AND
C--  THE OVERALL RISK FOR OFSS FOR PARTICULAR COMBINATION
C--   OF N1 ,N2 WITH N1+N2=N  .
C------------------------------------------------------
```

```
      DO 11 I=1,NM1
      R1=I-1
      NR1=N1-R1
      R4=L8-RD1-R1-1
      L6=N11-RD1-R1-1
      LR5=1-RD1-R1
      LL1=RD1+R1
      L7=LL1-1
      Q3=IV(L6)*IV(L7)*IV(L8)
      RQ3=1.0/Q3
      DO 11 L=1,NM2
      R2=L-1
      NR2=N2-R2
      LL2=RD2+R2
      SUM=0.0
      R5=1-RD2-R2
      LL6=N22-RD2-R2-1
      LR4=L8-RD2-R2-1
      LL7=LL2-1
      Q4=IV(LL6)*IV(LL7)*IV(L8)
      RQ4=1.0/Q4
      DO 29 J=LL2,LU2
      L1=LL1+J
      L2=R4-J
      L3=R5+J
      L4=J+1
      L5=R6-J
      Q5=IV(L4)*IV(L5)
      RQ5=1.0/Q5
   29 SUM=SUM+IV(L1)*IV(L2)*L3*RQ5
      S1=Q1*SUM*RQ3
      NW1=1
      SUM=0.0
      DO 14 J=LL1,LU1
      L1=LL2+J
      L2=LR4-J
      L3=LR5+J
      L4=J+1
      L5=LR6-J
      Q5=IV(L4)*IV(L5)
      RQ5=1.0/Q5
   14 SUM=SUM+IV(L1)*IV(L2)*L3*RQ5
      S2=Q2*SUM*RQ4
      IF(S2.GE.S1)GOTO 12
      S1=S2
12    P1=IV(N1)*IV(N2)*IV(L7)*IV(L6)
      P2=IV(LL7)*IV(LL6)*IV(ND11)*IV(ND22)
      P3=IV(R1)*IV(R2)*IV(NR1)*IV(NR2)
      P4=IV(LR6)*IV(R6)*IV(RD11)*IV(NRD1)*IV(RD22)*IV(NRD2)
      P=P1*P2/(P3*P4)
11    SUM1=SUM1+(P*S1)
      SUM1=SUM1+(N1*C1)+(N2*C2)
      IF(IM.EQ.1)GOTO 23
      IF(SUM1.GT.SUM2)GOTO 2
      IF(SUM1.LT.SUM2)GOTO 23
      SUM3=SUM1
      NOPT3=N1
```

```
      NOPT4=N2
      GOTO 2
 23   SUM2=SUM1
      NOPT1=N1
      NOPT2=N2
2     CONTINUE
      IF(SUM3.EQ.SUM2)GOTO 24
      WRITE(6,30)SUM2,NOPT1,NOPT2,N
      GOTO 1
 24   WRITE(6,30)SUM2,NOPT1,NOPT2,N
      WRITE(6,30)SUM3,NOPT3,NOPT4,N
1     CONTINUE
 30    FORMAT(5X,E14.6,10X,I3,5X,I3,5X,I3)
 20    FORMAT(5X,6I6)
      STOP
      END
```

Appendix (6.1)

Listing of Program OPST1

This program computes the risk values of the scheme OPT_1 under linear loss function and produces the exceptional points where OPT_1 and δ_1 have different decisions. The points and their types (whether S_1, S_2, B_1 or B_2) are saved in a file, say EPTS, to be used later. The number of these exceptional points is also computed and printed out.

The input data are:

(1) the value of δ_0 ($0 \leq \delta_0 \leq 1$); choose optimal value of δ_0; i.e. that value which gives minimum number of exceptional points,

(2) loss constants and sampling costs,

(3) prior information on p_1 and p_2,

(4) maximum total number of observations.

The main variable names used in the program (with the identifiers used in the text in brackets) are:

DEL	preassigned value (δ_0)
K1, K2	loss constants (K_1, K_2)
C1, C2	sampling costs (C_1, C_2)
RD1, ND1	prior number of successes, observations on population 1 (a_1, b_1)
RD2, ND2	prior number of successes, observations on population 2 (a_2, b_2)
R1, N1	actual number of successes, observations on population 1 (c_1, d_1)
R2, N2	actual number of successes, observations on population 2 (c_2, d_2)

Appendix (6.1) continued

S1, S2	stopping risks of taking decision D_1 and D_2 (S_1, S_2)
B1, B2	continuation risks with population 1 and 2 (B_1, B_2)
B	$\min(B_1, B_2)(B)$
S	array of optimal risks $(D(r_1, n_1, r_2, n_2))$
IND = 1, 2, 3, 4	type of risk $(S_1 \equiv 1, S_2 \equiv 2, B_2 \equiv 3, B_2 \equiv 4)$
N	maximum sample size (N).

```
      PROGRAM OPST1(INPUT,OUTPUT,RES,TAPE2=INPUT,
     1TAPE6=OUTPUT,TAPE4=RES)
      REAL  K1,K2,IV(0:150)
      INTEGER  RD1,RD2,R1,R2,R4,R5,R6
      ROWWISE S(128,11,11)
      COMMON S
       IV(0)=1.0
       DO 133 J=1,150
       IV(J)=IV(J-1)*J
 133   CONTINUE
      READ(2,*)DEL
      READ(2,*,END=10000)K1,K2,C1,C2
10000 READ(2,*,END=10001)N,ND1,ND2,RD1,RD2
10001 WRITE(6,30)K1,K2,C1,C2
      WRITE(6,55555)N,ND1,ND2,RD1,RD2
      WRITE(6,22222)DEL
22222 FORMAT(5X,'DELTA=',F2.1)
C     ------------------------------------------------------------
       NUM=0
       NUMI=0
      NM=N+1
C     ------------------------------------------------------------
C-- THE DO LOOP 11 COMPUTES THE STOPPING RISKS OF OPT1
C-- AND PERFORMS THE COMPARISON BETWEEN OPT1 AND
C-- THE SUBOPTIMAL SCHEME WHERE N=N1+N2
C-------------------------------------------------------------
      DO 11 IM=1,NM
      N2=IM-1
      N1=N-N2
      NM1=N1+1
      NM2=N2+1
      N11=ND1+N1
      N22=ND2+N2
      LU1=N11-1
      LU2=N22-1
      Q=IV(LU1)*IV(LU2)
      Q1=K1*Q
      Q2=K2*Q
      R6=N22-1
      LR6=N11-1
      L8=N11+N22-1
      DO 11 I=1,NM1
      R1=I-1
      R4=L8-RD1-R1-1
      L6=N11-RD1-R1-1
      LR5=1-RD1-R1
      LL1=RD1+R1
      L7=LL1-1
      Q3=IV(L6)*IV(L7)*IV(L8)
      RQ3=1.0/Q3
      T1=1.0
      T1=T1*LL1
      T1=T1/N11
      DO 11 L=1,NM2
      R2=L-1
      LL2=RD2+R2
      SUM=0.0
```

```
      R5=1-RD2-R2
      LL6=N22-RD2-R2-1
      LR4=L8-RD2-R2-1
      LL7=LL2-1
      Q4=IV(LL6)*IV(LL7)*IV(L8)
      RQ4=1.0/Q4
      T2=1.0
      T2=T2*LL2
      T2=T2/N22
      DO 29 J=LL2,LU2
      L1=LL1+J
      L2=R4-J
      L3=R5+J
      L4=J+1
      L5=R6-J
      Q5=IV(L4)*IV(L5)
      RQ5=1.0/Q5
   29 SUM=SUM+IV(L1)*IV(L2)*L3*RQ5
      S1=Q1*SUM*RQ3
      SUM=0.0
      DO 14 J=LL1,LU1
      L1=LL2+J
      L2=LR4-J
      L3=LR5+J
      L4=J+1
      L5=LR6-J
      Q5=IV(L4)*IV(L5)
      RQ5=1.0/Q5
   14 SUM=SUM+IV(L1)*IV(L2)*L3*RQ5
      S2=Q2*SUM*RQ4
      IF(S1.LE.S2)THEN
        IND=1
      ELSE
        S1=S2
        IND=2
      ENDIF
      IF(T1.LE.T2)THEN
        IF(IND.EQ.1)THEN
          NUMI=NUMI+1
        ELSE
          IF(IND.EQ.2)THEN
            NUM=NUM+1
            WRITE(4,139)R1,N1,R2,N2,IND
          ENDIF
        ENDIF
      ELSE
        IF(IND.EQ.1)THEN
           WRITE(4,139)R1,N1,R2,N2,IND
           NUM=NUM+1
        ELSE
           IF(IND.EQ.2)THEN
             NUMI=NUMI+1
            ENDIF
        ENDIF
      ENDIF
 11   S(IM,I,L)=S1
C-----------------------------------------------------------
```

```
C  ---THE DO LOOP 12 COMPUTES THE STOPPING AND CONTINUATION
C  ---RISKS OF OPT1 AND PERFORMS COMPARISON BETWEEN OPT1
C  ---AND THE SUBOPTIMAL SCHEME WHERE N1+N2=N-1 TO 0
C----------------------------------------------------
      DO 12 IN=1,N
      NN=N-IN
       NM=NN+1
      DO 13 IM=1,NM
      N2=IM-1
      N1=N-IN-N2
      N11=ND1+N1
      N22=ND2+N2
       RN11=1.0/(N11)
       RN22=1.0/(N22)
      NM1=N1+1
      NM2=N2+1
      LU1=N11-1
      LU2=N22-1
      Q=IV(LU1)*IV(LU2)
      Q1=K1*Q
      Q2=K2*Q
      R6=N22-1
      LR6=N11-1
      L8=N11+N22-1
      DO 17 I=1,NM1
      R1=I-1
      R4=L8-RD1-R1-1
      L6=N11-RD1-R1-1
      LL1=RD1+R1
      L7=LL1-1
      LR5=1-RD1-R1
      Q3=IV(L6)*IV(L7)*IV(L8)
      RQ3=1.0/Q3
      T1=1.0
      T1=T1*LL1
      T1=T1/N11
      DO 17 L=1,NM2
      R2=L-1
      LR4=L8-RD2-R2-1
      LL6=N22-RD2-R2-1
      LL2=RD2+R2
      T2=1.0
      T2=T2*LL2
      T2=T2/N22
      LL7=LL2-1
      R5=1-RD2-R2
       Q4=IV(LL6)*IV(LL7)*IV(L8)
       RQ4=1.0/Q4
      B1=C1+((RD1+R1)*S(IM,I+1,L)+
     1(N11-RD1-R1)*S(IM,I,L))*RN11
      B=B1
      B2=C2+((RD2+R2)*S(IM+1,I,L+1)+(N22-RD2-R2)
     1*S(IM+1,I,L))*RN22
      IF(B1.GT.B2)THEN
      B=B2
      ENDIF
       SUM=0.
```

```
      DO 28 J=LL2,LU2
       L1=LL1+J
       L2=R4-J
        L3=R5+J
       L4=J+1
       L5=R6-J
       Q5=IV(L4)*IV(L5)
       RQ5=1.0/Q5
   28  SUM=SUM+IV(L1)*IV(L2)*L3*RQ5
       S1=Q1*SUM*RQ3
       SUM=0.0
       DO 19 J=LL1,LU1
       L1=LL2+J
       L2=LR4-J
       L3=LR5+J
       L4=J+1
       L5=LR6-J
       Q5=IV(L4)*IV(L5)
       RQ5=1.0/Q5
  19     SUM=SUM+IV(L1)*IV(L2)*L3*RQ5
       S2=Q2*SUM*RQ4
      IF(S2.GE.S1)GOTO 22
      S(IM,I,L)=S2
      IND=2
      GOTO 77
22    S(IM,I,L)=S1
      IND=1
77    IF(S(IM,I,L).LE.B)GOTO 33
      S(IM,I,L)=B
      IF(B1.LE.B2)THEN
        IND=3
      ELSE
        IND=4
      ENDIF
33    T=T1-T2
      IF(ABS(T).LE.DEL)GOTO 36
      IF(T1.LE.T2)GOTO 46
      IF(T1.GT.T2)THEN
        IF(IND.EQ.1)THEN
          WRITE(4,139)R1,N1,R2,N2,IND
          NUM=NUM+1
        ELSE
          IF(IND.EQ.2)THEN
            NUMI=NUMI+1
          ELSE
            IF(IND.EQ.3)THEN
              WRITE(4,139)R1,N1,R2,N2,IND
              NUM=NUM+1
            ELSE
              IF(IND.EQ.4)THEN
                WRITE(4,139)R1,N1,R2,N2,IND
                NUM=NUM+1
              ENDIF
            ENDIF
          ENDIF
        ENDIF
      ENDIF
```

```
      GOTO 17
46    IF(IND.EQ.1)THEN
        NUMI=NUMI+1
      ELSE
        IF(IND.EQ.2)THEN
          WRITE(4,139)R1,N1,R2,N2,IND
          NUM=NUM+1
        ELSE
          IF(IND.EQ.3)THEN
            WRITE(4,139)R1,N1,R2,N2,IND
            NUM=NUM+1
          ELSE
            IF(IND.EQ.4)THEN
              WRITE(4,139)R1,N1,R2,N2,IND
              NUM=NUM+1
            ENDIF
          ENDIF
        ENDIF
      ENDIF
      GOTO 17
36    IF(T1.LT.T2)GOTO 20
      IF(T1.GT.T2)GOTO 39
      IF(N11.GT.N22)GOTO 20
39    IF(IND.EQ.1)THEN
        WRITE(4,139)R1,N1,R2,N2,IND
        NUM=NUM+1
      ELSE
        IF(IND.EQ.2)THEN
          WRITE(4,139)R1,N1,R2,N2,IND
          NUM=NUM+1
        ELSE
          IF(IND.EQ.3)THEN
            NUMI=NUMI+1
          ELSE
            IF(IND.EQ.4)THEN
              WRITE(4,139)R1,N1,R2,N2,IND
               NUM=NUM+1
             ENDIF
           ENDIF
         ENDIF
      ENDIF
      GOTO 17
20    IF(IND.EQ.1)THEN
        WRITE(4,139)R1,N1,R2,N2,IND
        NUM=NUM+1
      ELSE
        IF(IND.EQ.2)THEN
          WRITE(4,139)R1,N1,R2,N2,IND
          NUM=NUM+1
        ELSE
          IF(IND.EQ.3)THEN
            WRITE(4,139)R1,N1,R2,N2,IND
            NUM=NUM+1
           ELSE
            IF(IND.EQ.4)THEN
              NUMI=NUMI+1
             ENDIF
```

```
          ENDIF
        ENDIF
      ENDIF
   17  CONTINUE
    13 CONTINUE
C-------------------------------------------------------
    12 CONTINUE
       WRITE(6,21)NUM
       WRITE(6,23)NUMI
21     FORMAT(5X,'NO.OF OPT.PTS.NOT AGREE WITH SUBOPT=',I10)
23     FORMAT(5X,'NO.OF OPT.PTS.AGREE WITH SUBOPT=',I10)
55555 FORMAT(BZ,5I5)
    30 FORMAT(BZ,8E14.6)
  139   FORMAT(5I3)
       STOP
       END
```

Appendix (6.2)

Listing of Program OPST2

The routine SORMRG, available in CDC7600 operating system at UMRCC, is capable of doing three functions: Sort only, Merge only and Sort and Merge. Our purpose from using this routine is to sort the exceptional set of points (EPTS), produced by OPST1, in ascending order so that all points starting with particular integer value (0-9) in the first place (first dimension) will be put in a group and the ordering will take place with each group in the same way.

The output data is SEPTS.

PROGRAM OPST2

```
LIBRARY(PROCLIB)
NEWPROC.
GETFEP(ST,OSMR10)
FILE(IN1,RT=F,FL=15)
COPY(ST,IN1)
REWIND(IN1)
SORTMRG.
REWIND(OUT1)
COPY(OUT1,D)
PUTFEP(D,SOSMR10)
##S
SORT
FILE,INPUT=IN1,OUTPUT=OUT1
FIELD,NUMA(1,3,DISPLAY),NUMB(4,3,DISPLAY),NUMC(7,3,DISPLAY)
,NUMD(10,3,DISPLAY),NUME(13,3,DISPLAY)
KEY,NUMA(A,ASCII6),NUMB(A,ASCII6),NUMC(A,ASCII6),NUMD(A,ASCII6)
,NUME(A,ASCII6)
END
```

Appendix (6.3)

Listing of Program OPST3

This program carries out the simulation for OPT_1 with generated p_1 and p_2 from uniform distribution using the subroutine GΦ5CAF of NAG Library, available at UMRCC.

The input data are:

(1) the exceptional set of points SEPTS and the number of them,

(2) the same optimal value of δ_0 used in Program OPST1,

(3) the same prior information on p_1 and p_2 used in Program OPST1,

(4) the number of runs.

The main variable names used in the program (with identifiers used in the text in brackets) are:

DELTA	preassigned value (δ_0)
RD1, ND1	prior number of successes, observations on population 1 (a_1, b_1)
RD2, ND2	prior number of successes, observations on population 2 (a_2, b_2)
N	the maximum number of observations (N)
R1, N1	actual number of successes, observations on population 1 (c_1, d_1)
R2, N2	actual number of successes, observations on population 2 (c_2, d_2)
T1, T2	the posterior estimates of p_1, p_2 (r_1/n_1, r_2/n_2).

The program has the advantage of spreading the number of runs (t) over several times of executing the program through the read statement

```
READ(2, *) IFLAG, INCR
```

and adjusting

Appendix (6.3) continued

MMLOW = 1 + INCR

MMUPP = MMLOW + 9999

where,

IFLAG be zero for starting from first run or non-zero for continuation

INCR be the increase in the number of runs in each time the program executed

MMLOW, MMUPP be the range of the number of runs where the first is the lower value and the second is the upper value.

Example:

(1) If we wish to use t = 10000 runs then we input

IFLAG = ϕ and INCR = ϕ;

consequently MMLOW = 1 and MMUPP = 10000.

(2) If we wish to use t = 10000 runs with a program which is a time consuming then we can, for example execute the program with 1000 runs at each time as follows

(a) At the beginning, let IFLAG = ϕ and INCR = ϕ and change 9999 to 999. The program then will perform the first 1000 runs since MMLOW = 1 and MMUPP = 1000. The results will be saved in a file and restored in the next execution of the program.

(b) In the second execution, let IFLAG $\neq \phi$ and INCR = 1000. The program will restore the results saved in stage (a) and continue calculations for this second 1000 runs which starts from MMLOW = 1001 to MMUPP = 2000. The new results, based on 2000 runs, will again be saved to restore in the next execution.

(c) Repeat step (b) until t = 10000.

```
      PROGRAM OPST3(INPUT,OUTPUT,DAT3,DAT4,DAT5,TAPE2=INPUT,
     1TAPE6=OUTPUT,TAPE3=DAT3,TAPE5=DAT5,TAPE4=DAT4)
      REAL X1,X2,P1,P2,DELTA,T1,T2,T
      INTEGER ND1,ND2,RD1,RD2,R1,R2,N1,N2
       INTEGER RT,R11,N11,R22,N22
      LOGICAL TEST
      DIMENSION IB(9),XA(4)
      COMMON IP1PT(11),IP2(259),IP3(259),IP4(259),IP5(259)
      COMMON /REST/ DELTA,ESMAL,ESS,ESUCC,I,IA,IB,II,
     1              ILOW,IP1,IUPP,J,MRU,MRUN,MS1,
     2              MS2,N,ND1,ND2,NPT,NR,NRU,NRUN,NSMAL,NS1,NS2,
     3              NT,NUM,N1,N11,N2,N22,PCS,PCS1,PCS2,PERROR,
     4              PERR1,PERR2,PNOT,PNUT,P1,P2,RD1,RD2,RT,R1,R11,
     5              R2,R22,T,TEST,T1,T2,XA,X1,X2,X3,X4
C         ALL VARIABLES IN THE PROGRAM PUT INTO THE 'REST'
C         COMMON BLOCK - ALL EXCEPT IP1PT,IP2,IP3 AND IP4,IP5
C         THE COMMON BLOCK IS EQIVALENCED TO RESTAR
      DIMENSION RESTAR(126)
      EQUIVALENCE (RESTAR(1),DELTA)
C     ..FUNCTION REFERENCES..
      REAL G05CAF
C     ..SUBROUTINE REFERENCES..
C     G05CBF
C     ..
C      NS1 NO OF TIMES TO STOP AND TAKE D2/D1 IN MM RUNS
C      NS2 NO OF TIMES TO STOP AND TAKE D1/D2 IN MM RUNS
C      NT TOTAL NO OF OBSERVATIONS TO STOP IN MM RUNS
C      PERR1=P(D1/D2) IN MM RUNS
C      PERR2=P(D2/D1) IN MM RUNS
C      PERROR=PERR1+PERR2,THE TOTAL PROB. OF ERROR IN MM RUNS
C      ESS=NT/MM,THE EXPECTED S S IN MM RUNS
C      NSMAL=NO.OF OBSERVATIONS ON SMALLER PROB.POPS IN MM RUNS
C      ESMAL=NSMAL/MM,THE EXPECTED NO.OF OBSERVATIONS ON
C      SMALLER PROB.POP.
C      RT=NO. OF SUCCESSES FROM POP1 AND POP2 IN MM RUNS
C      ESUCC=RT/MM,THE EXPECTED NO OF SUCCESSES IN MM RUNS
C       NUM NO OF POINTS NOT FOLLOW SUBOPT SCHEME
C       DELTA SOME NONNEGATIVE VALUE BETWEEN 0 AND 1
C-----------------------------------------------------------------------
C       READ THE START/CONTINUATION FLAG
      READ(2,*) IFLAG,INCR
C       IFLAG IS (0/NOT 0) FOR (START/CONTINUE) RUN
      IF (IFLAG.EQ.0) THEN
       WRITE(6,18887)
18887 FORMAT(5X,'START RUN')
      READ(2,*)NUM
      WRITE(6,99996)NUM
      II=-1
      DO 4 IA=1,NUM
      READ(3,*)IP1,IP2(IA),IP3(IA),IP4(IA),IP5(IA)
      IF(IP1.EQ.II)GOTO 4
      I=IP1+1
      IP1PT(I)=IA
      II=IP1
    4 CONTINUE
      IP1PT(IP1+2)=NUM+1
      WRITE(6,90010)IP1PT
```

```
      READ(2,*)DELTA
      READ(2,*)N,ND1,ND2,RD1,RD2
      WRITE(6,99999)
      WRITE(6,99998)DELTA
      WRITE(6,99997)N,ND1,ND2,RD1,RD2
      WRITE(6,99994)N
C--G05CBF SETS THE BASIC GENERATOR ROUTINE G05CAF
C       TO A REPEATABLE INITIAL STATE-----
      CALL G05CBF(0)
      NS1=0
      NS2=0
      MS1=0
      MS2=0
      PCS=0.0
      NT=0
      PERROR=0.0
      ESS=0.0
      NSMAL=0
      RT=0
      NRUN=0
      MRUN=0
      PNUT=0.0
           ELSE
88999 WRITE(6,18886)
18886 FORMAT(5X,'CONTINUATION')
C--G05CGF RESTORES THE STATE OF THE BASIC GENERATOR
C        ROUTINE G05CAF AFTER A PREVIOUS CALL TO G05CFF---
      READ(5) IP1PT,IP2,IP3,IP4,IP5,RESTAR
      WRITE(6,99997)NS1,NS2,MS1,MS2,NT
      WRITE(6,99997)RT,NSMAL,NRUN,MRUN,NPT
      WRITE(6,99999)
      WRITE(6,99998)DELTA
      WRITE(6,99997)N,ND1,ND2,RD1,RD2
      CALL G05CGF(IB,9,XA,4,1)
           ENDIF
      MMLOW=1+INCR
      MMUPP = MMLOW+9999
       WRITE(6,99997)MMLOW,MMUPP
      DO 11, J=MMLOW,MMUPP
C--G05CAF RETURNS A PSEUDO-RANDOM REAL NUMBER TAKEN
C     FROM A UNIFORM DISTRIBUTED BETWEEN 0 AND 1---
      X3=G05CAF(X3)
      P1=X3
      X4=G05CAF(X4)
      P2=X4
      NR=N+1
      R1=0
      N1=0
      R2=0
      N2=0
      NRU=0
       MRU=0
      DO 10 I=1,NR
      R11=RD1+R1
      N11=ND1+N1
      R22=RD2+R2
      N22=ND2+N2
```

```
      T1=FLOAT(R11)/FLOAT(N11)
      T2=FLOAT(R22)/FLOAT(N22)
      T=T1-T2
      IF(I.EQ.NR)GOTO 17
      IF(ABS(T).LE.DELTA)GOTO 15
   17   IF(T1.GT.T2)GOTO 24
      IND=1
      TEST=.FALSE.
      ILOW=IP1PT(R1+1)
      IUPP=IP1PT(R1+2)-1
      DO 40 IA=ILOW,IUPP
      IF((IP2(IA).EQ.N1).AND.(IP3(IA).EQ.R2
     1).AND.(IP4(IA).EQ.N2))TEST=.TRUE.
      IF(TEST)GOTO 22
   40 CONTINUE
       MRU=MRU+1
      GOTO 23
   22 NRU=NRU+1
      IF(IP5(IA).EQ.2)THEN
        GOTO 14
      ELSE
        IF(IP5(IA).EQ.3)THEN
          GOTO 6
        ELSE
          IF(IP5(IA).EQ.4)THEN
            GOTO 5
          ENDIF
        ENDIF
      ENDIF
      GOTO 12
   23 IF(P1.LE.P2)GOTO 77
      NS2=NS2+1
      GOTO 12
 24   IND=2
      TEST=.FALSE.
      ILOW=IP1PT(R1+1)
      IUPP=IP1PT(R1+2)-1
      DO 42 IA=ILOW,IUPP
      IF((IP2(IA).EQ.N1).AND.(IP3(IA).EQ.R2
     1).AND.(IP4(IA).EQ.N2))TEST=.TRUE.
      IF(TEST)GOTO 27
   42 CONTINUE
      MRU=MRU+1
      GOTO 14
   27 NRU=NRU+1
      IF(IP5(IA).EQ.1)THEN
        GOTO 23
      ELSE
        IF(IP5(IA).EQ.3)THEN
          GOTO 6
        ELSE
          IF(IP5(IA).EQ.4)THEN
            GOTO 5
          ENDIF
        ENDIF
      ENDIF
      GOTO 12
```

```
   14 IF(P1.GT.P2)GOTO 88
      NS1=NS1+1
      GOTO 12
  15  IF(T1.LT.T2)GOTO 20
      IF(T1.GT.T2)GOTO 19
      IF(N11.GT.N22)GOTO 20
19    IND=3
      TEST=.FALSE.
      ILOW=IP1PT(R1+1)
      IUPP=IP1PT(R1+2)-1
      DO 50 IA=ILOW,IUPP
      IF((IP2(IA).EQ.N1).AND.(IP3(IA).EQ.
     1R2).AND.(IP4(IA).EQ.N2))TEST=.TRUE.
      IF(TEST)GOTO 7
   50 CONTINUE
      MRU=MRU+1
      GOTO 6
    7 NRU=NRU+1
      IF(IP5(IA).EQ.1)THEN
        GOTO 23
       ELSE
        IF(IP5(IA).EQ.2)THEN
          GOTO 14
        ELSE
          IF(IP5(IA).EQ.4)THEN
             GOTO 5
           ENDIF
        ENDIF
      ENDIF
      GOTO 12
6     NT=NT+1
      X1=G05CAF(X1)
      IF(X1.GT.P1)GOTO 35
      R1=R1+1
   35 N1=N1+1
      GOTO 10
20    IND=4
       TEST=.FALSE.
      ILOW=IP1PT(R1+1)
      IUPP=IP1PT(R1+2)-1
      DO 60 IA=ILOW,IUPP
      IF((IP2(IA).EQ.N1).AND.(IP3(IA).EQ.R2
     1).AND.(IP4(IA).EQ.N2))TEST=.TRUE.
      IF(TEST)GOTO 8
   60 CONTINUE
      MRU=MRU+1
      GOTO 5
    8 NRU=NRU+1
      IF(IP5(IA).EQ.1)THEN
        GOTO 23
      ELSE
        IF(IP5(IA).EQ.2)THEN
          GOTO 14
        ELSE
          IF(IP5(IA).EQ.3)THEN
            GOTO 6
          ENDIF
```

```
        ENDIF
      ENDIF
      GOTO 12
 5    NT=NT+1
      X2=G05CAF(X2)
      IF(X2.GT.P2)GOTO 45
      R2=R2+1
   45 N2=N2+1
   10 CONTINUE
   77 MS1=MS1+1
      GOTO 12
   88 MS2=MS2+1
   12 NRUN=NRUN+NRU
      MRUN=MRUN+MRU
      RT=RT+R1+R2
      IF(P1.LE.P2)GOTO  33
      NSMAL=NSMAL+N2
      GOTO 11
   33  NSMAL=NSMAL+N1
   11 CONTINUE
      NPT=NRUN+MRUN
C--G05CFF SAVES THE CURRENT STATE OF THE BASIC GENERATOR
C    ROUTINE G05CAF-----
       CALL G05CFF(IB,9,XA,4,1)
      WRITE(4) IP1PT,IP2,IP3,IP4,IP5,RESTAR
      PCS1=FLOAT(MS1)/FLOAT(MMUPP)
      PCS2=FLOAT(MS2)/FLOAT(MMUPP)
      PCS=PCS1+PCS2
      WRITE(6,99997)NS1,NS2,MS1,MS2,NT
99996 FORMAT(5X,I8)
      WRITE(6,99997)RT,NSMAL,NRUN,MRUN,NPT
      WRITE(6,88888)PCS
88888 FORMAT(5X,'PCS=',F10.4)
      PERR1=FLOAT(NS2)/FLOAT(MMUPP)
      PERR2=FLOAT(NS1)/FLOAT(MMUPP)
      PERROR=PERR1+PERR2
      ESS=FLOAT(NT)/FLOAT(MMUPP)
      ESMAL=FLOAT(NSMAL)/FLOAT(MMUPP)
      ESUCC=FLOAT(RT)/FLOAT(MMUPP)
      PNOT=FLOAT(NRUN)/FLOAT(NPT)
      ERTN=(ESUCC*N)/ESS
      WRITE(6,99993)PERR1
      WRITE(6,99992)PERR2
      WRITE(6,99989)PERROR
      WRITE(6,99991)ESS
       WRITE(6,99985)ESMAL
       WRITE(6,99984)ESUCC
      WRITE(6,99983)PNOT
      WRITE(6,99982)ERTN
       STOP
99999 FORMAT(5X,'SIMULATION RESULTS FOR OPT1')
99998 FORMAT(5X,F10.4)
99997 FORMAT(5X,6I6)
99994 FORMAT(5X,'SAMPLE SIZE=',I5)
99993 FORMAT(5X,'PERR1=P(D1/D2)=',F10.4)
99992 FORMAT(5X,'PERR2=P(D2/D1)=',F10.4)
99991 FORMAT(5X,'EXP.S.S=TOT.NO.OF.OBS.TO.STOP/NO.OF.RUNS=',F10.4)
```

```
99989 FORMAT(5X,'TOTAL PROBABILITY OF ERROR=',F10.4)
99985 FORMAT(5X,'E(NO. OF OBSERVATIONS.SMALL PROB.POP.)=',F10.4)
99984 FORMAT(5X,'E(R1+R2)=',F10.4)
99983 FORMAT(5X,'P(DOESNOT FOLLOW SUBOPT DESIGN)=',F10.4)
99982 FORMAT(5X,'E(RETURN)=',F10.4)
90010 FORMAT(1X,I10)
      END
```

Appendix (9.1)

Listing of Program MULPE

This program produces simulation results for the multinomial selection scheme under the stopping rule MS_1 (δ_0). The cell probabilities are generated using Method ME_2.

The input data are:

(1) prior information on the cell probabilities,

(2) the number of runs.

The main variable names used in the program (with the identifiers used in the text in brackets) are:

NC	the number of cells $(k + 1)$
CD	array of prior cell frequencies $(n'_1, n'_2, \ldots, n'_{k+1})$
MM	the number of runs (t)
DEL	preassigned value (δ_0)
C	array of actual cell frequencies $(n_1, n_2, \ldots, n_{k+1})$
CDD	array of posterior cell frequencies $(n''_1, n''_2, \ldots, n''_{k+1})$
P	array of cell probabilities $(p_1, p_2, \ldots, p_{k+1})$
PHAT	array of posterior cell probabilities $(\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_{k+1})$
NT	actual total number of observations (m)
N	maximum total number of observations (N)
CTOT	total posterior number of observations $(m' + m)$.

```
       PROGRAM MULPE(INPUT,OUTPUT,TAPE2=INPUT,TAPE6=OUTPUT)
      INTEGER C,CD,CDD,CKK,CTOT
      REAL LLT,LP
      DIMENSION IND1(11),IND2(11),Q(11),Z(11),INX(11),
     1 IPL(11),IPLH(11)
      COMMON/OUTPT/PCST,PERRT,ESS,NC,MM,N,CD(11),NSIZE,PST(11)
      COMMON/OUTPT1/DELV(11),PCSTV(11),PERRTV(11),ESSV(11)
       COMMON/OUTPT2/PCOV(11)
      DIMENSION PCS(11),PERR(11),CDD(11),C(11),P(11),PHAT(11)
       DIMENSION CKK(11)
C--     FUNCTION REFRENCES--
      REAL G05CAF
C-----------------------------------------------------
C
      READ(2,*)NI,MM,NC,(CD(I),I=1,NC)
       DO 1 NJ=1,NI
       N=NJ*10
       DO 3 NSIZE=1,1
       NF=N/NSIZE
       DO 2 LD=1,1
       DEL=(0.1*LD)+0.2
      CALL G05CBF(0)
      DO 22 I=1,NC
      IPL(I)=0
      PCS(I)=0.0
      PERR(I)=0.0
 22   IPLH(I)=0
      PCST=0.0
      PERRT=0.0
      NNT=0
      NTIME=0
C
       ICO=0
C THE DO LOOP 11 GIVE NO. OF RUNS
C
      DO 11 J=1,MM
      ZTOT=0.0
      DO 96 I1=1,NC
       G=FLOAT(CD(I1))
      Z(I1)=G05DGF(G,1.0,0)
      ZTOT=ZTOT+Z(I1)
 96   CONTINUE
      DO 97 I1=1,NC
      P(I1)=Z(I1)/ZTOT
 97   CONTINUE
      DO 9 I1=1,NC
      Q(I1)=P(I1)
 9    CONTINUE
      DO 5 II=1,NC
  5   IND1(II)=II
      DO 4 L=1,NC-1
      M=L+1
      DO 4 K=M,NC
      IF(Q(L).GE.Q(K))GOTO 4
      AK=Q(L)
      Q(L)=Q(K)
      Q(K)=AK
```

```
      K1=IND1(L)
      IND1(L)=IND1(K)
      IND1(K)=K1
 4    CONTINUE
      KC=0
      DO 99 K2=2,NC
      IF(Q(1).EQ.Q(K2))GOTO 100
        GOTO 110
 100  KC=KC+1
  99  CONTINUE
 110  IF(KC.EQ.0)GOTO 120
      ULT=0.0
      LLT=0.0
      Y=G05CAF(Y)
      VAL=FLOAT(1)/FLOAT(KC+1)
      DO 130 K3=1,KC+1
       ULT=ULT+VAL
       LLT=ULT-VAL
       IF(Y.GT.LLT.AND.Y.LT.ULT)THEN
        IND1(1)=IND1(K3)
        ENDIF
 130   CONTINUE
 120  NT=0
      NR=(NF)+1
      DO 12 I1=1,NC
  12  C(I1)=0
C
C
      DO 10 I=1,NR
      CTOT=0
      DO 19 L1=1,NC
      CKK(L1)=CD(L1)+C(L1)
       CDD(L1)=CKK(L1)
  19  CTOT=CTOT+CKK(L1)
      DO 14 L2=1,NC
      PHAT(L2)=FLOAT(CDD(L2))/FLOAT(CTOT)
  14  CONTINUE
      DO 6 II=1,NC
  6   IND2(II)=II
      DO 20 L=1,NC-1
       M=L+1
       DO 20 K=M,NC
        IF(PHAT(L).GE.PHAT(K))GOTO 20
      AK=PHAT(L)
      PHAT(L)=PHAT(K)
      PHAT(K)=AK
      K1=IND2(L)
      IND2(L)=IND2(K)
      IND2(K)=K1
  20  CONTINUE
      IC=0
      DO 44 K2=2,NC
      IF(PHAT(1).EQ.PHAT(K2))GOTO 24
       GOTO 23
  24  IC=IC+1
  44  CONTINUE
  23  IF(IC.EQ.0)GOTO 67
```

```
      ULT=0.0
      LLT=0.0
      Y=G05CAF(Y)
      VAL=FLOAT(1)/FLOAT(IC+1)
      DO 26 K3=1,IC+1
      ULT=ULT+VAL
      LLT=ULT-VAL
      IF(Y.GT.LLT.AND.Y.LT.ULT)THEN
       IND2(1)=IND2(K3)
       ENDIF
 26   CONTINUE
C
C-------STOPPING RULE-----
C
 67   IF(NT.EQ.N)GOTO 50
       PDIF=PHAT(1)-PHAT(2)
       IF(PDIF.GE.DEL)GOTO 50
C
C-----SAMPLING RULE------
C
       DO 101 NG=1,NSIZE
      LP=0.0
      UP=0.0
       NT=NT+1
       Y=G05CAF(Y)
      DO 15 L=1,NC
      UP=UP+P(L)
      LP=UP-P(L)
      IF(Y.GT.LP.AND.Y.LT.UP)GOTO 16
      GOTO 15
  16  C(L)=C(L)+1
       GOTO 101
 15   CONTINUE
 101   CONTINUE
 10    CONTINUE
C
C------DECISION RULE-----
C
 50   DO 31 L1=1,NC
 31   INX(L1)=L1
      DO 35 JJ1=1,NC
      IF(IND2(1).EQ.INX(JJ1))THEN
        IF(Q(1).EQ.P(JJ1))THEN
          IPL(JJ1)=IPL(JJ1)+1
          GOTO 113
         ELSE
          IPLH(JJ1)=IPLH(JJ1)+1
          GOTO 113
         ENDIF
       ENDIF
 35    CONTINUE
 113    DO 36 JJ2=1,NC
        IF(IND2(JJ2).EQ.IND1(JJ2))GOTO 36
         GOTO 13
 36     CONTINUE
        ICO=ICO+1
 13   IF(NT.LT.N)THEN
```

```
      NTIME=NTIME+1
      ENDIF
      NNT=NNT+NT
 11   CONTINUE
C
C THE END OF DO LOOP 11
C
      DO 33 II1=1,NC
      PCS(II1)=FLOAT(IPL(II1))/FLOAT(MM)
 33   PERR(II1)=FLOAT(IPLH(II1))/FLOAT(MM)
      DO 34 NN1=1,NC
      PCST=PCST+PCS(NN1)
      PERRT=PERRT+PERR(NN1)
 34   CONTINUE
      ESS=FLOAT(NNT)/FLOAT(MM)
      PSTOP=FLOAT(NTIME)/FLOAT(MM)
       PCO=FLOAT(ICO)/FLOAT(MM)
      DELV(LD)=DEL
      PCSTV(LD)=PCST
      PERRTV(LD)=PERRT
      ESSV(LD)=ESS
       PST(LD)=PSTOP
       PCOV(LD)=PCO
  2   CONTINUE
      CALL PRINTR
  3   CONTINUE
 1    CONTINUE
      STOP
      END
      SUBROUTINE PRINTR
       INTEGER CD
      COMMON/OUTPT/PCST,PERRT,ESS,NC,MM,N,CD(11),NSIZE,PST(11)
      COMMON/OUTPT1/DELV(11),PCSTV(11),PERRTV(11),ESSV(11)
      COMMON/OUTPT2/PCOV(11)
      CHARACTER*60,TITLE
      CHARACTER*30,LAB1,LAB2,LAB3,LAB4,LAB5
      CHARACTER*10,DE,PC,PER,EN,PS,PO,SPA,DAS
      DAS='----------'
      SPA='          '
      DE=' DELTA     '
      PC='     PCST  '
      PER='  P(ERROR)'
      EN='     E(N) '
       PS='     PSTOP'
       PO='     PCO'
      TITLE='     SIMULATION RESULTS FOR PE-GROUP(IND)'
      LAB1='      GROUP SIZE='
      LAB2='      NO OF  RUNS='
      LAB3='      PRIOR PROBS='
      LAB4='      GENERATED P-VALUES'
      LAB5='      SAMPLE SIZE='
      WRITE(6,'(A)')TITLE
      WRITE(6,'(A)')SPA
      WRITE(6,'(A,I6)')LAB2,MM
      WRITE(6,'(A,6I3)')LAB3,NC,(CD(L5),L5=1,NC)
      WRITE(6,'(A)')LAB4
      WRITE(6,'(A,I6)')LAB5,N
```

```
      WRITE(6,'(A,I6)')LAB1,NSIZE
      WRITE(6,'(A)')SPA
      WRITE(6,'(6A)')DAS,DAS,DAS,DAS,DAS,DAS
      WRITE(6,'(6A)')DE,PC,PER,EN,PS,PO
      WRITE(6,'(6A)')DAS,DAS,DAS,DAS,DAS,DAS
      WRITE(6,30)(DELV(I),PCSTV(I),PERRTV(I),ESSV(I),PST(I)
     1,PCOV(I),I= 1,1)
      WRITE(6,'(6A)')DAS,DAS,DAS,DAS,DAS,DAS
      WRITE(6,'(A)')SPA
      WRITE(6,'(A)')SPA
      WRITE(6,'(A)')SPA
30    FORMAT(F5.2,6X,5F10.4)
      RETURN
      END
```